

Open Research Online

The Open University's repository of research publications and other research outputs

The Uncanny Valley Effect

Thesis

How to cite:

Lay, Stephanie Claire (2015). The Uncanny Valley Effect. PhD thesis The Open University.

For guidance on citations see [FAQs](#).

© 2015 The Author



<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Version: Version of Record

Link(s) to article on publisher's website:

<http://dx.doi.org/doi:10.21954/ou.ro.0000a94c>

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

The Uncanny Valley Effect

Thesis submitted for the
Degree of Doctor of Philosophy
by
Stephanie Lay

BA (Hons) Psychology - University of Sussex
MSc Psychology - Open University
MSc Psychological Research Methods - Open University

Department of Psychology
Faculty of Social Sciences
The Open University, UK

2015

Abstract

The Uncanny Valley Effect (UVE) first emerged as a warning against making industrial robots appear so highly human-like that they could unsettle the real humans around them. It proposed a specific pattern of negative emotional responses to entities that were almost but not quite human, and has been proposed as the reason why some entities such as dolls, mannequins and zombies may appear unsettling.

The aim of this thesis was to move beyond an anecdotal explanation to understand more about the perception of near-human faces, and how this compares to the perception of human and non-human faces. The aims were to explore the relationship between the human-likeness of faces and emotional responses to them, to understand reactions to and descriptions of near-human faces, to explore aspects of how near-human faces are processed and to explore whether mismatched emotional expressions might contribute to the perception of some near-human faces as eerie.

Five studies were carried out using face images whose human-likeness was systematically controlled or measured. A non-linear relationship between human-likeness and eeriness was found, but the near-human faces were not always the eeriest images. Near-human faces were found to be subject to the effects of inversion, and inversion was found to heighten perceptions of eeriness. Faces were created which contained mismatched emotional expressions, and the blends combining happy faces with angry or fearful eyes were rated as the most eerie. Incongruities between aspects of appearance or behaviour had been cited as explanations for the UVE in the past but this thesis presents the first evidence that differences in eeriness may result from incongruities between emotional expressions. Directions for future research have been suggested to explore these findings in a wider context and to understand more about the UVE.

Declaration

This thesis consists of my own original work and comprises less than 100,000 words.

All material used within this thesis has been acknowledged where appropriate.

Acknowledgements

I would like to extend my sincere thanks to Dr Nicky Brace, Professor Graham Pike and Professor Frank Pollick for supervising my research. Thank you for being so generous with your support and expertise but also for sharing your enthusiasm for research and the joy of discovery.

I would like to warmly acknowledge the contribution made by everyone who has taken part in my studies over the years, as without your interest and engagement I would not have a thesis to share. Thank you.

I could not have completed this thesis without the support of my husband Jef, and the encouragement and inspiration from my parents Shirley and Steve and my sister Sophie. Thank you so much for your patience while I've had my attention elsewhere. This thesis is dedicated to you all - the best family anyone could wish for.

Contents

| | |
|---|----|
| Abstract | 1 |
| Declaration | 2 |
| Acknowledgements | 2 |
| Contents | 3 |
| List of tables..... | 8 |
| List of figures | 10 |
| Chapter 1: Introduction and Literature Review | 15 |
| 1.1: Locating The Uncanny Valley Effect..... | 15 |
| 1.1.1 Defining the Uncanny Valley Effect | 15 |
| 1.1.2 Early 20 th Century Views of the UVE..... | 20 |
| 1.1.3 Establishing Research Into the UVE | 23 |
| 1.2: Proposed Explanations for the UVE..... | 29 |
| 1.2.1 Evolutionary Aesthetics | 29 |
| 1.2.2 Category Boundaries and Categorical Perception | 31 |
| 1.2.3 Neuropsychological Perspectives | 32 |
| 1.2.4 Fears of Own Mortality..... | 35 |
| 1.3: Empirical Tests of the UVE..... | 37 |
| 1.3.1 Evidence of Evolutionary Aesthetics..... | 37 |
| 1.3.2 Evidence of Category Boundaries and Categorical Perception..... | 40 |
| 1.3.3 Neuropsychological Evidence | 42 |
| 1.3.4 Trust, Social Actions and Empathy..... | 44 |
| 1.3.5 Error Sensitivity..... | 51 |
| 1.3.6 Anomalous Features | 53 |

| | |
|---|-----|
| 1.3.7 Cue Mismatch Effects | 57 |
| 1.4: Summary and Review of UVE Literature | 66 |
| 1.5: Face Perception Literature | 67 |
| 1.4.1 Introducing Analytic and Holistic Processing | 68 |
| 1.4.2 Developmental approaches | 71 |
| 1.4.3 FIE as a loss of configural representation | 73 |
| 1.4.4 Alternative explanations for the FIE | 77 |
| 1.6: Emotional Expression Literature..... | 81 |
| 1.5.1 Processing expressions of emotion | 81 |
| 1.5.2 Experimental Studies of Emotional Expression | 83 |
| 1.5.3 Concealment and deception: the effects of partially-seen expressions..... | 87 |
| 1.5.4 Expression and Inversion | 88 |
| 1.7: Research Design | 92 |
| Chapter 2: Methodology | 93 |
| 2.1: Theoretical Background..... | 93 |
| 2.1.1 Structure of the Methodology Chapter | 93 |
| 2.1.2 Theoretical Background..... | 93 |
| 2.2: General Methodological Issues | 95 |
| 2.2.1 Research Approach..... | 95 |
| 2.2.2: Survey Research..... | 96 |
| 2.2.3: Collecting Data Online | 98 |
| 2.2.4: Coding Qualitative Data..... | 100 |
| 2.2.5: Controlled Experimental Study..... | 102 |
| 2.3: Methodological Issues Pertaining to the Uncanny Valley | 104 |
| 2.3.1: The Circularity Problem | 104 |

| | |
|--|-----|
| 2.3.2: Operationalising the UV graph | 105 |
| 2.3.3: Near-humans as research stimuli | 110 |
| 2.4 Ethical Considerations | 112 |
| 2.4.1: Ethical Clearance | 112 |
| 2.4.2: General Considerations | 113 |
| 2.4.3: UVE-Specific Considerations..... | 115 |
| Chapter 3: Phase 1: Exploring The Valley | 117 |
| 3.1: Introduction to Phase 1 | 117 |
| 3.2: Study 1: Subjective ratings of near-human agents..... | 122 |
| 3.2.1: Introduction to Subjective ratings of near-human agents..... | 122 |
| 3.2.2: Method | 124 |
| 3.2.3: Results | 125 |
| 3.2.4: Discussion | 128 |
| 3.3: Study 2: Imagined Encounters with near-human agents..... | 132 |
| 3.3.1: Introduction to Imagined encounters with near-human agents | 132 |
| 3.3.2: Method | 143 |
| 3.3.3: What will the subjective ratings of human-likeness, strangeness and eeriness be for an identified collection of NHAs? | 151 |
| 3.3.4: When asked to describe the faces of five NHAs, which features will be emphasised and how will this compare to the descriptions of human and non-human comparison faces?..... | 157 |
| 3.3.5: When considering a hypothetical social encounter with the NHAs, what terms will be used to express how people would feel about this and which emotions will characterise their reactions? | 164 |

| | |
|--|-----|
| 3.3.6: How will the ratings of eeriness relate to the identified features and emotions?..... | 170 |
| 3.4: General Discussion of Phase 1..... | 173 |
| Chapter 4: Phase 2: The Inversion Effect..... | 180 |
| 4.1: Introduction to Phase 2 | 180 |
| 4.2: Study 3: Ratings of morph faces for inversion study | 183 |
| 4.2.1: Introduction to Ratings of morph faces..... | 183 |
| 4.2.2: Method | 186 |
| 4.2.3: Results | 194 |
| 4.2.4: Discussion | 209 |
| 4.3: Study 4: The effect of inversion on recognising near-human faces..... | 213 |
| 4.3.1: Introduction to Effect of inversion | 213 |
| 4.3.2: Method | 216 |
| 4.3.3: Results | 221 |
| 4.3.4: Discussion | 226 |
| 4.4: General Discussion of Phase 2..... | 228 |
| Chapter 5: Phase 3: Study 5: Facial Expressions of Emotion | 234 |
| 5.1: Introduction to Phase 3 | 234 |
| 5.2: Method | 244 |
| 5.3: Results and Discussions | 254 |
| 5.3.1: Eeriness ratings for different face image types | 254 |
| 5.3.2: Emotional responses to face image types | 273 |
| 5.3.3: Emotion identification | 288 |
| 5.3.4: Emotion classification..... | 297 |
| 5.4: General Discussion of Phase 3..... | 300 |

| | |
|--|-----|
| Chapter 6: General Discussion..... | 306 |
| 6.1: Overview of Research Findings..... | 306 |
| 6.2: Implications for the Uncanny Valley Theory..... | 320 |
| 6.3: Future Directions | 323 |
| 6.4: Reflections on the Uncanny Valley Effect..... | 325 |
| References | 327 |
| Appendices | 335 |
| Phase 1 | 335 |
| Appendix 1a: Participant demographics for the imagined encounters study | 335 |
| Appendix 1b: Example survey webpages | 336 |
| Appendix 1c: Frequency cross-tabulations of individual positive, negative and neutral emotion terms by face type..... | 342 |
| Phase 2 | 348 |
| Appendix 2a: Participant demographics for rating study | 348 |
| Appendix 2b: Participant demographics for inversion study..... | 348 |
| Appendix 2c: Image sources | 349 |
| Appendix 2d: Foil and experimental images..... | 353 |
| Appendix 2e: Online study texts..... | 354 |
| Appendix 2f: Participant Information and consent forms for face to face study | 359 |
| Appendix 2g: Screenshots from face to face study | 362 |
| Phase 3 | 367 |
| Appendix 3a: Participant demographic tables..... | 367 |
| Appendix 3b: Survey web page examples | 369 |
| All Phases..... | 374 |
| Appendix 4: Eeriness Ratings across all Phases | 374 |

List of tables

| | |
|--|-----|
| Table 1: Research principles for investigating the UVE | 119 |
| Table 2: Overall means and standard deviations for ratings of human-likeness, strangeness and eeriness | 126 |
| Table 3: Mean human-likeness, strangeness and eeriness by image..... | 126 |
| Table 4: The three images (11, 7 and 1) rated as medium-high human-likeness with high strangeness or eeriness..... | 131 |
| Table 5: Mean frequencies of references to key terms for five face types..... | 158 |
| Table 6: Mean frequencies of references to key emotion terms for five face types. | 165 |
| Table 7: Mean references to terms synonymous with eeriness for each face type. | 168 |
| Table 8: Summary of correlation results for H8, H9 and H10 | 171 |
| Table 9: Verbatim comments where references were made to eyes in a creepy or unsettling context | 176 |
| Table 10: Increases in human-likeness between morph stages. | 196 |
| Table 11: Mean response times and standard deviations for each morph stage, upright and inverted. | 223 |
| Table 12: Structure of tasks, IVs, DVs, measures and survey questions. | 247 |
| Table 13: Results of post-hoc t-tests on mean fear ratings..... | 276 |
| Table 14: Results of post-hoc t-tests on mean sadness ratings | 278 |
| Table 15: Results of post-hoc t-tests on mean disgust ratings..... | 280 |
| Table 16: Results of post-hoc t-tests on mean happiness ratings..... | 281 |
| Table 17: Summary of Pearson's correlation tests on mean ratings of eeriness and fear, sadness, disgust, anger, happiness and surprise..... | 283 |
| Table 18: Mean fear, anger, disgust, happiness and sadness for the two eeriest blends. | 287 |

| | |
|--|-----|
| Table 19: Results of post-hoc t-tests on mean emotion identification accuracy | 290 |
| Table 20: Distribution of participant ages | 335 |
| Table 21: Distribution of participant gender | 335 |
| Table 22: Participant location | 335 |
| Table 23: Frequencies of key positive emotion terms for five face types, by descending frequency. | 343 |
| Table 24: Frequencies of key negative emotion terms for five face types, by descending frequency. | 345 |
| Table 25: Frequencies of key neutral emotion terms for five face types, by descending frequency. | 345 |
| Table 26: Negative terms used to in descriptions of emotional reactions towards near- human agents, indicating whether they are synonymous with eeriness..... | 347 |
| Table 27: Participant demographics for rating study | 348 |
| Table 28: Participant demographics for inversion study | 349 |
| Table 29: Original image source links for Phase 2 | 352 |
| Table 30: All foil and experimental images used for Phase 2, by image category and non- human/human anchor. | 354 |
| Table 31: Participant demographics for Phase 3..... | 369 |
| Table 32: Table comparing mean eeriness ratings for images used in each research phase. | 375 |

List of figures

| | |
|---|-----|
| Figure 1: The Uncanny Valley. (MacDorman (2005a), after Mori (1970).)..... | 16 |
| Figure 2: Potential outcomes from a preferential looking experiment with three levels of realism. (From Steckenfinger and Ghazanfar (2009)). | 38 |
| Figure 3: Tinwell (2009)'s graph showing three distinct valleys, which occurred when plotting mean human-likeness against familiarity. | 61 |
| Figure 4: Non-face objects which are readily seen as face-like: from Little et al (2011).... | 69 |
| Figure 5: Scatter plot of mean human-likeness and strangeness: data labels refer to picture references. | 127 |
| Figure 6: Scatter plot of mean human-likeness and eeriness: data labels refer to picture references. | 128 |
| Figure 7: An extract of a diagram taken from A.D. Milner and M. Goodale (1995), The Visual Brain In Action, demonstrating the dorsal and ventral streams of visual processing in the macaque monkey brain. Reproduced by permission of Oxford University Press. (From Breen (2000)) | 136 |
| Figure 8: Summary of research findings relating lost and preserved abilities to the dual pathway hypothesis. | 137 |
| Figure 9: Five face types presented to participants in this study..... | 146 |
| Figure 10: Comparison of the mean rating scales for each face type. | 151 |
| Figure 11: Plot of mean human-likeness against strangeness, overlaid with the curve that would be expected in an UVE..... | 154 |
| Figure 12: Holistic terms were most frequently used when describing the Human and Cluster 3 faces, analytic terms for the Cluster 2 face. | 159 |

| | |
|--|-----|
| Figure 13: Mean number of references to eyes used when describing the five different face types. | 160 |
| Figure 14: Bar chart comparing mean frequencies of references to key terms for five face types. | 165 |
| Figure 15: Image set for one example of a statue to human morph. | 187 |
| Figure 16: Image set for one example from each image category. | 187 |
| Figure 17: Screenshot of the Fantamorph software package, showing the set-up of correspondence dots for two faces. | 190 |
| Figure 18: Mean human-likeness with error bars for each image category. | 195 |
| Figure 19: Mean human-likeness with error bars for each morph stage. | 196 |
| Figure 20: A comparison of human-likeness ratings for four image categories as the morphed faces progressed from non-human to human | 198 |
| Figure 21: Mean strangeness with error bars for each image category. | 199 |
| Figure 22: Mean strangeness with error bars for each morph stage. | 200 |
| Figure 23: A comparison of strangeness ratings for four image categories as the morphed faces progressed from non-human to human | 202 |
| Figure 24: Mean eeriness with standard error bars for each image category. | 203 |
| Figure 25: Mean eeriness with error bars for each morph stage. | 204 |
| Figure 26: A comparison of eeriness ratings for four image categories as the morphed faces progressed from non-human to human | 205 |
| Figure 27: Mean human-likeness and strangeness for animal faces. | 207 |
| Figure 28: Mean human-likeness and strangeness for doll faces. | 208 |
| Figure 29: Mean human-likeness and strangeness for robot faces. | 208 |
| Figure 30: Mean human-likeness and strangeness for statue faces. | 209 |

| | |
|---|-----|
| Figure 31: Test image set for an animal to human morph at the 75% human, 25% artificial point. | 214 |
| Figure 32: Inverted and upright versions of the original and testing images. | 217 |
| Figure 33: Images which were not correctly identified by participants in Study 3..... | 222 |
| Figure 34: Mean response time for each image category. | 225 |
| Figure 35: Mean response time for each morph stage. | 225 |
| Figure 36: Three examples of face expression blends, from left to right: <i>happy face, neutral eyes; sad face, happy eyes; happy face, fearful eyes</i> | 238 |
| Figure 37: Examples of all 24 levels of FEBs showing how the expression images were combined..... | 246 |
| Figure 38: Diagram showing the survey flow and elements encountered by participants. | 251 |
| Figure 39: Comparison of the mean eeriness for matched and mismatched blends when presented upright and inverted. | 256 |
| Figure 40: Heat map visualising eeriness ratings for each individual face image, presented upright and inverted..... | 258 |
| Figure 41: Examples of the three face images where fearful eyes were presented, either blended with a happy or neutral face or in the unblended fearful face..... | 259 |
| Figure 42: Comparison of mean eeriness for fearful-eyed faces when presented upright and inverted. | 260 |
| Figure 43: Examples of the three face images where angry eyes were presented, either blended with a happy or neutral face or in the unblended angry face. | 261 |
| Figure 44: Comparison of mean eeriness for angry-eyed faces when presented upright and inverted. | 262 |
| Figure 45: Examples of the positive, negative and neutral face image types..... | 264 |

| | |
|--|-----|
| Figure 46: Mean eeriness ratings for positive, negative and neutral blend types, by orientation..... | 265 |
| Figure 47: Comparing mean eeriness ratings for face image types when presented upright. | 266 |
| Figure 48: Heat map visualising eeriness ratings for the five neutral-eyed blends. | 269 |
| Figure 49: Face blends sorted by descending eeriness when inverted, highlighted to present difference between inverted and upright eeriness ratings. | 271 |
| Figure 50: Mean emotion strength across all face blends, presented upright and inverted. Tables are grouped by whether the blend was matched or mismatched, and sorted in descending order of the emotion strength when presented upright..... | 274 |
| Figure 51: Comparison of the mean fear strength for matched and mismatched blends when presented upright and inverted..... | 276 |
| Figure 52: Comparison of the mean sadness strength for matched and mismatched face images when presented upright and inverted. | 278 |
| Figure 53: Comparison of the mean disgust strength for matched and mismatched face images when presented upright and inverted. | 279 |
| Figure 54: Comparison of the mean happiness strength for matched and mismatched face images when presented upright and inverted. | 281 |
| Figure 55: Eeriness ratings plotted against mean fear, sadness, disgust, anger, happiness and surprise..... | 283 |
| Figure 56: Comparison of the mean % accuracy for matched and mismatched face images when presented upright and inverted. | 289 |
| Figure 57: Heat map presenting the emotions identified for each blend, presented upright, indicating whether this was a correct or incorrect identification..... | 291 |

| | |
|---|-----|
| Figure 58: Heat map presenting the emotions identified for each blend, presented inverted, indicating whether this was a correct or incorrect identification..... | 291 |
| Figure 59: Comparison of correctly identified emotions by whether these were identified in the eyes or the surrounding face, when presented upright or inverted..... | 294 |
| Figure 60: Comparison of happy/non happy emotions identified in blended faces where happy faces were combined with different types of non-happy eyes. | 295 |
| Figure 61: Categorisation ratings for the blends presented..... | 298 |
| Figure 62: Comparing 'cannot decide' and other categories by blend. | 299 |
| Figure 63: Eeriness ratings of face images presented in each research phase, indicating those rated as most and least eerie. | 309 |

Chapter 1: Introduction and Literature Review

1.1: Locating The Uncanny Valley Effect

1.1.1 Defining the Uncanny Valley Effect

More recently, androids and computers have taken on the attributes of living beings. As many examples from literature and film demonstrate, however, there remains a profound divide between artificial life and our own. [...] Paradoxically, when looking at a dead body or a dead animal, the human brain 'reads' it as related to the living - we know it is dead but it is from life. By contrast, however 'lifelike' a robot or avatar, the brain does not (or cannot) perceive it as alive. What is still lacking in the collision between the organic and the technological is the 'experience' of a digital face, or a bodily or sensory response to the image. Without these, the digitised expressive elements fail to communicate. The face is powerful in this respect: as an index of the heart, it has so far proved impossible to replicate digitally. (Kemp (2004), p140-141.)

The uncanny valley effect (UVE) was once a little-known idea relating to robotics design and how people's reactions to artificial agents changed as they were made to look increasingly human-like. It related to changes in emotional responses to mechanical objects as they begin to take on human characteristics and they initially become more appealing and likeable as human-likeness increases. However, this only holds true up to a certain point as when those human characteristics were sufficiently realistic to look almost but not quite human, the reaction changed and rather than continuing to increase in acceptability, the agent suddenly appeared eerie and seemed unsettling. This sudden dip in emotional response is the 'valley' component of the term. It was termed an 'uncanny' valley as the responses to those near-human agents (NHAs) or near-human entities (NHEs) were characterised by unease or disquiet rather than simply being rejected without an affective component. The term was originally coined by Mori (1970) although Reichardt (1978) is credited as the first person to use the term in English. Mori's use of the term in the original Japanese was *bukimi no tani* which was translated as meaning 'valley of eeriness'. Agents cited as falling into the uncanny valley (UV) include

computer game characters where appearance and behaviour are mismatched, department store mannequins with realistic faces but blank eyes and animated characters where the synchronised voices are just slightly mis-timed. The UVE is best described with reference to the following graph:

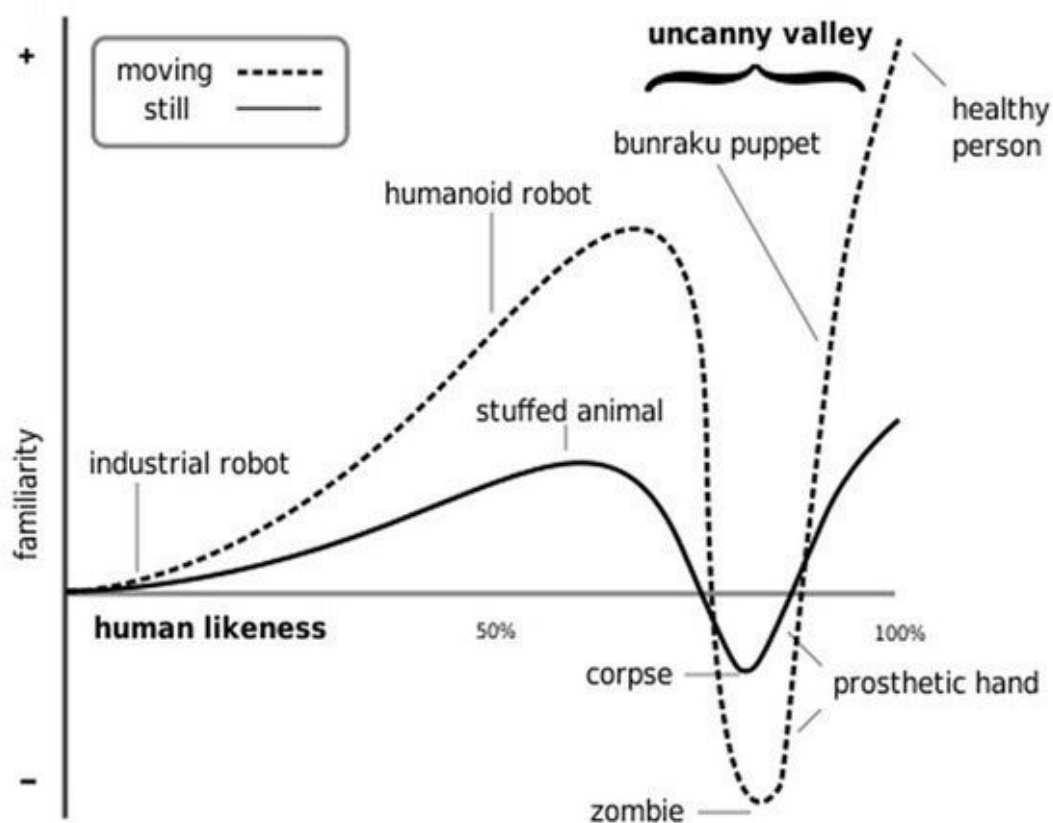


Figure 1: The Uncanny Valley. (MacDorman (2005a), after Mori (1970).)

This representation of the valley is based on an idea described by Mori (1970), which described a hypothetical relationship between ‘human-likeness’ and ‘familiarity’, defining the human-likeness dimension using examples of entities ranging from the extremely un-human-like (an industrial robot) to completely human (a healthy living person). The ‘familiarity’ dimension is more complex to define, particularly given its dual meaning in English as it can indicate either an absence of novelty or a sense of closeness so this axis has been interpreted variously in terms of positive affect, increasing affinity and emotional warmth. The graph depicts dual curves, describing that the hypothesised

relationship between human-likeness and familiarity also varies as a function of whether the agent is moving or still. Both curves initially increase in familiarity as the human-likeness increases until the 60-65% human point where familiarity begins to decrease, finally reaching its lowest point at around 75-80% human. After this point, it rises steeply again until the highest familiarity is reached for a moving and completely human-like agent (100% human-like could either be a real human, or something so close that it is no longer possible to distinguish it from one.) The curves vary in magnitude according to whether the entities are still or moving, but on both there is a distinct dip in familiarity between 75% and 90% human where familiarity plummets, forming the valley in the term UV.

The UV concept originated from Mori's observations of how people reacted when non-human entities acquired some human qualities and began to become more human-like. To illustrate the idea he offers a striking example of how encountering a realistic prosthetic hand can feel:

'Some prosthetic hands attempt to simulate veins, muscles, tendons, finger nails, and finger prints, and their color resembles human pigmentation. So maybe the prosthetic arm has achieved a degree of human verisimilitude on par with false teeth. But this kind of prosthetic hand is too real and when we notice it is prosthetic, we have a sense of strangeness. So if we shake the hand, we are surprised by the lack of soft tissue and cold temperature. In this case, there is no longer a sense of familiarity. It is uncanny. In mathematical terms, strangeness can be represented by negative familiarity, so the prosthetic hand is at the bottom of the valley. So in this case, the appearance is quite human like, but the familiarity is negative. This is the uncanny valley.' ('The Uncanny Valley, Appendix B', in MacDorman (2005a).)

Mori gives three examples to illustrate what might fall into the valley, starting with zombies, which are well-recognised as unsettling, and also adding corpses and prosthetic hands which may be potentially disquieting if encountered when unexpected. Mori did not offer additional examples but it could be seen that clowns, masks and dolls designed to be particularly realistic could also live in the same valley. The original article was a

short paper in Japanese, and was written for an audience of roboticists and engineers. It posed questions about the nature of human-likeness and suggested that designers working in the field of robotics should pay careful attention to the methods they used to make their creations appealing or acceptable to the public, challenging the prevalent goal of aiming to make robots as human as possible. It was translated into English by MacDorman (2005a) and began to gain popular attention, particularly as Mori revisited the theory twice more in his writings. Mori (2005) reflected on the idea that when someone dies, their lack of animation may be unsettling but it may also suggest that the person is more at peace if the death released them from suffering or uncertainty, and this peaceful aspect may moderate any sense of uncanniness. Mori (2005) also suggested that there may be entities that deserve to be positioned at the highest form of acceptability on the curve and proposed that the faces of some Buddhist statues may actually appear more elegant, calm and dignified as idealised portrayals of the human form than genuine humans can be, so it may have been wrong to position human beings as the highest point on the original curve. Mori (2012) was an updated re-translation of the 1970 original which resolved an ambiguity with the term of 'familiarity' as used in the original paper. This re-translation relabelled the 'familiarity' axis as 'affinity' as this is a clearer term than familiarity which could refer to levels of novelty or strangeness. The 2012 paper suggested that the UVE is caused by the viewer discovering the 'deception' that the agent is not actually as human as it appears, and it is the discovery of a deceptively human appearance that causes eeriness when the initial affinity drops away.

The UVE was presented as a theoretical construct in 1970, and at the time was proposed as a thought experiment rather than an explanatory framework. The classic graph was based on Mori's subjective judgments of the familiarity and human-likeness of each of the examples included in the graph and the curves linking them are illustrative

rather than based on any empirical findings. Human-likeness was presented as a percentage but the placement of the examples seems arbitrary, as a stuffed animal is arguably less human-like than a humanoid robot rather than more as it appears on the graph. However, the basic premise is a tantalising one for exploration: is there any validity to the idea that almost-human entities have unsettling properties and if so, is it possible to identify attributes of those entities that evoke a sense of uncanniness in a reliable and psychologically grounded way?

The UVE concept lay dormant for several years before it caught the attention of animators, psychologists and philosophers early in the 21st century. The discipline has progressed well beyond its initial status as a thought experiment and now receives serious consideration as a topic of psychological interest in its own right as well as being relevant to the domains of human-computer interaction and android design. There has been an increase in the number of studies which have provided experimental evidence that there is a pattern of responses to NHAs which matches the shape described in Mori's graph: for example the studies by MacDorman (2006), Seyama and Nagayama (2009), Steckenfinger and Ghazanfar (2009), and Tinwell et al (2011b) which will be described in detail later in this review. This thesis has been developed since 2006, starting from the premise that research at that time was mainly focused on demonstrating that an UVE existed, and that there was a lack of psychological evidence for why the effect might occur for certain types of near-human or virtual agent, nor was there a detailed description of what set the UVE apart from other types of emotional response to a stimulus. The present research was designed to address that knowledge gap by attempting to systematically evoke an UVE under controlled conditions in order to draw reliable conclusions about the types of stimuli that do (or do not) elicit a UVE response. Over a series of studies, psychological research techniques were used to draw conclusions

regarding how NHEs are perceived and to contribute to the overall sum of knowledge about the UVE.

1.1.2 Early 20th Century Views of the UVE

The UVE was only translated into English in 2005 but the concept of uncanniness as a more general term was the subject of two psychoanalytic papers in the early 1900s. They formed a background of psychological thinking about uncanniness as a term in its own right, and that was distinct and separate from the more general emotions of fear or disgust. These two papers are therefore useful for setting a background and context for approaches to exploring the UVE.

Jentsch's (1906) paper 'On the Psychology of the Uncanny' asked what can cause a feeling of uncanniness and gave examples that could elicit uncanniness regularly, powerfully and generally in everyday life, particularly how it could be used in literature to elicit a reaction from a reader. He suggested that a sense of uncanniness occurred when an entity or event deviates from what is expected and everyday, but is more than a simple rejection of novelty and involves the uncertainty of not being aware whether the entity or event is quite what it appears to be. The core of the uncanny sense was seen as a feeling of being lost, out of place or disoriented which can be gleaned from the literal meaning of the word in German where 'unheimlich' translates as 'unhomely', meaning that that which is homely and familiar is comfortable and acceptable, but that which is unhomely is unfamiliar and unsettling. Doubt over whether an entity is living or dead is given as one example as suddenly realising that one's assumption of animacy is wrong could be profoundly unsettling. For instance, it is easy to imagine being frightened by waxworks in a darkened museum where something inanimate is momentarily and alarmingly perceived as animate. This doubt about the animacy of an object is described

as key to eliciting a sense of uncanniness as it is a failure to understand and integrate sensory input from the world leading to one of two unsettling outcomes. Wrong conclusions could be drawn about the nature of an object, leading to an unsettling reaction when these conclusions are revealed as inaccurate, or it may simply be impossible to be sure about the fundamental nature of an entity, which is in itself an unpleasant sensation. Jentsch predicted that children, the intoxicated or those of a nervous or sensitive disposition might be the most susceptible to uncanniness and also that darkness is particularly effective at evoking an uncanny response, as is the uncertainty that can arise from a very noisy environment, such as a factory floor, which could cause distractions where decisions about the true nature of things may be particularly difficult.

Freud (1919) wrote about the uncanny as a special subset of that which is frightening and used Jentsch's literary and everyday observations to introduce a psychoanalytic description of the sensation. He agreed that the sense of the uncanny was related to novelty and unfamiliarity but that there was more to the effect than just the fear of new things. He approached the subject firstly through a lengthy linguistic analysis of how the word can be translated in several different languages and concluded that a simplistic and direct translation is rarely possible but that most versions include facets of the gruesome, evil or disgusting as well as the sense of disorientation and unfamiliarity conveyed in the German word. The paper proceeds with examples of 'things, persons, impressions, events and situations' that evoke the uncanny. He proposed four factors which serve to distinguish a frightening experience from an uncanny one: firstly, in considering 'animism, magic and sorcery and the omnipotence of thoughts' he suggested that one source of the uncanny is the inexplicable or the poorly understood which includes a component of a primitive understanding of the world where all entities contained spirits. An uncanny

experience is an evocation of this less enlightened interpretation of the world. Secondly, Freud suggested that our attitudes to death and particularly our own mortality can trigger a sense of the uncanny particularly where entities evoke doubt as to whether they are alive or dead. Third, uncanniness occurs in 'involuntary repetition' which is described with reference to perceiving elements of your own appearance in other people or unintentionally catching sight of one's own reflection under circumstances where it is not immediately clear that one is looking at oneself. The final explanation suggests that a sense of the uncanny is actually a manifestation of the 'castration complex' and explained with reference a story introduced in Jentsch's paper about a 'Sandman'. Jentsch had cited it as an example where a writer uses uncertainty about whether an entity is animate or not to elicit unease. Freud explains that the entity in question is a doll who presents a convincingly and unsettlingly human appearance. While the intellectual uncertainty of whether the doll is human or not seems undoubtedly disturbing, he notes that there is a more direct source of unease in the story where the protagonist has a childhood fear of his eyes being stolen by the mythical 'Sandman' of the title. Freud's interpretation of this fear is not just the natural horror of mutilation and blinding but that it evokes (in males) a childhood fear of castration. In this way, the uncanny sense is the recurrence of a repressed fear from childhood as an uncomfortable intrusion into adult life.

The present research is not conducted in the psychoanalytic tradition and as such the notion of a castration complex is not a relevant framework for exploration, but the other examples presented by Freud and Jentsch do merit consideration when looking at the uncanny in relation to almost-human entities. Both authors concluded that the closer something is to human, the more likely it is to spark a sense of uncanniness which is at the core of Mori's proposal of the UVE. Jentsch's article allows a framework for considering the uncanny from a psychological perspective as it locates uncanniness as an

unpleasant feeling related to fear and aversion, often occurring when the nature of something is uncertain, and which exhibits individual differences in how susceptible people are to experiencing it. However, neither study provides empirical support for what makes an entity uncanny. To understand more about how researchers began to address that question, this review will now consider studies which move on from uncanniness in general to looking in detail at Mori's idea of an UVE.

1.1.3 Establishing Research Into the UVE

Research into the UVE began to proliferate from 2005, following a themed workshop ('Views of the Uncanny Valley, 2005') and an UVE-focused special edition of the 'Interaction Studies' journal in 2006 which developed several of the theories presented at the workshop. Four themes emerged from these early explorations.

The first theme was around how the nature of the 'familiarity' axis can be best understood. Pollick (2005) examined how the importance of how the movement and appearance of an animated entity are integrated and suggested that a lack of such integration may be responsible for the UVE. A technique using gesture exaggeration was adopted by extrapolating positions within a movement space and finding prototypical locations that aid recognition. The perception of emotion as conveyed through bodily movement and facial expressions was examined using computer generated (CG) models where neutral movements and expressions were matched to emotional movements and expression and the ability of participants to correctly recognise these was measured. Identification performance varied by emotion, with anger being read mainly from movement and happiness and sadness from the facial expression. While these observations offer useful insights for the perception of form and motion, Pollick concluded that at that stage the UV was a descriptive rather than predictive term. To

move beyond this and investigate the valley fully, it was suggested that both axes of the graph needed to be better understood, but that attention should be focused particularly on the dimension of 'familiarity' where the nature of the emotional response in the effect required greater definition. Brenton et al (2005) questioned whether the UVE could be linked to the concept of 'presence' in virtual reality. To achieve a sense of 'being there' in a virtual reality environment that feels comparable to being in a real world environment, participants make and test a series of hypotheses about their surroundings. Virtual reality settings which offer poor or confusing input to the viewer make it hard to test those hypothesis which causes an unsettling 'break in presence' (BIP) to occur by breaking the illusion and revealing that the virtual world is not real. Due to the unsettling qualities of a BIP, Brenton et al suggested that BIP is to presence what the UV is to human-likeness, and that it would be appropriate to employ the same investigative methods used for BIP in exploring the UV such as questionnaires (Slater et al, 2003) and measurements of skin conductance response and heart rate (Slater, 2002). Ziemke and Lindblom (2006) argued that it may be functionally impossible to evaluate whether the responses to androids occur because participants perceive them as non-human as opposed to novel or suspicious and raised concerns that there was a significant measurement issue in evaluating responses, especially if responses to the almost-human occur at a subconscious level as these could not be measured using verbal reports of conscious perceptions and would be difficult to measure using physiological measures of arousal. Hanson et al (2005) challenged whether the UVE was even a problem at all, as a survey of responses to videos of two human-like robots found that reported that 73% of respondents found both of the robots appealing and none reported them as disturbing. There were also no reports of the robots looking 'dead' as opposed to 'lively'. He also presented still images from a morphed sequence in which a cartoon figure is transformed

into a human figure. Participants were asked to rank whether they found the images acceptable or unacceptable. Hanson reported the median acceptability and argued that as none of the images were rated below 70%, where 100% indicated completely acceptable, this negated the existence of a UV. However, this finding is questionable. Firstly, the dimension used to rank the images (acceptability) is different from Mori's 'familiarity' scale. Secondly, although the acceptability does not dip particularly low, the fourth image of six, which is almost but not quite human, does have the lowest acceptability rating and indeed to counter the UV idea there should be no such negative peak at that point. This paper does not, therefore, constitute sufficient reason to abandon the UV concept.

The second theme introduced the role of one's own mortality in triggering perceptions of the uncanny. MacDorman (2005b) asked whether human-like robots are eerie because they evoke memories of death and dead bodies, and used images taken from a morph sequence between a robot and a human. Participants were asked to identify the eeriest point in the sequence as well as to provide ratings of several CG agents and robots. The eeriest agent was then used in an experiment designed to measure the extent to which participants associated ideas of death and mortality with humanoid robots. The researchers concluded that the participants who saw the eeriest of the two robots were more likely to exhibit behaviours relating to managing their fear of death. This cognitive explanation of the UV suggested further avenues for research to establish whether it could be elicited more generally across a range of near human entities.

The third theme was the role of category boundaries: Ramey (2005) suggested that the UVE was the result of difficulties in making decisions across a particular set of categorical boundaries. He argued that the UVE is not related solely to robots or the

almost-human but extends to many other instances where it is difficult to make boundary distinctions, such as knowing when a heap of sand stops being a heap of sand when single grains are removed one at a time. When the first grain is removed, the heap retains its heap status. With repeated removals, a point will be reached where that status is no longer valid - it is no longer a heap. Ramey asked at what point can that distinction be made, and suggested that the UVE is what happens when it is no longer possible to make that distinction, not between heap and not-heap but between non-human and human. There are other examples where this distinction problem may occur such as someone going bald one hair at a time, or the development of a foetus from a single cell but he observed that those transitions don't evoke a sense of eeriness at the category boundary and wondered about the special qualities of human-likeness that could cause a UVE to emerge.

The final theme considered neuropsychological approaches to the UVE. Berthouze and Berthouze (2005) used a mixed methods study to examine responses to the affective postures of avatars. Participants were asked to identify emotions from low-level descriptions of avatar appearance and the results indicated that they were particularly accurate at identifying the arousal of the avatars, but made more errors when identifying the valence of the emotion or the motion tendency. An fMRI measured the cortical activity when participants were asked to evaluate the expressiveness of an avatar and the face fusiform gyrus and anterior cingulate were identified as potential areas of activity when looking at near-human avatars. Chaminade et al (2005) used a motor interference paradigm to explore how seeing different types of artificial agent performing a particular action would interfere with participant's ability to perform that action themselves. Three levels of human-likeness were presented: a mechanical-looking robot, a humanoid robot and a human. In support of the UV hypothesis, Chaminade et al found that the humanoid

robot did cause this type of interference but that the human and mechanical robot did not. In a second study, animated models were presented at differing levels of detail with the lowest level of detail (dots identifying joints moving through space) acquiring more human-like traits until the final model depicted a person jogging, presenting the most detailed representation. Participants classified whether the movement was natural or artificial and sensitivity and bias towards biological motion were identified as measures. A UVE was found in both measures as sensitivity increased as the models became more human-like until the penultimate model where it dipped before peaking with the jogger model. Biological response bias, operating in the opposite direction, peaked with the penultimate model and was lowest for the lowest level representation. This second study was repeated using fMRI to examine the brain regions involved in perceiving different levels of human-like motion and the team reported that the right anterior insula responded strongly to perceiving the full model but not as strongly to the lower levels of representation.

Keysers and Gazzola (2005) used fMRI studies to explore the role of mirror circuits in the perception of humanoid robots. They argued that as computers replace humans in many areas of life, it will be important to understand how people might react when encountering a NHA and used the theory of 'shared circuits' which are activated when someone is observing an action performed by another person as well as performing it themselves. Such a shared circuit would be extremely efficient for empathetic communication because it would negate the need to describe one's state if it was possible to mirror the state of someone else. Keysers and Gazzola asked whether such shared circuits could be activated by watching a robot perform an action, in the same way that they would be activated if a human performed it. They found that shared circuits were robustly activated even if the robots were not particularly human-like: as long as

their behaviours had some human resonance, the shared circuitry was still activated. They also explored participant responses to being touched or viewing touch and somatosensory cortical activity was measured to look for the presence of secondary activity when participants saw but did not experience touch, and whether that still occurred if it was an inanimate object experiencing the touch. Keysers and Gazzola did indeed find that secondary activity was comparable regardless of whether a person or object was being touched, suggesting that our ability to take the perspective of others is not restricted to people. Their conclusion that humanoid robots are able to evoke mechanisms often thought to be purely human and social in origin is broadly supportive of the UVE as it suggests we are able to respond to these NHAs in a human-like way. However, none of the neuropsychological studies linked their findings to whether participants experienced any sense of unease when encountering the NHAs, so while their findings do suggest response patterns that can be explored in more detail, at that stage there was no attempt to link uncanniness to those responses.

It can be seen that by 2006, android science was becoming a burgeoning area for enquiry and that research into the UV had been suggested from a range of different perspectives but no explanations were yet forthcoming. Gee et al (2005) observed that too many studies were small-scale and observational in nature where responses to individual androids were used to draw conclusions about the UVE. Applying findings based on the specific qualities of one robot would be hard to generalise even to all other robots, let alone to the range of NHAs that are created in the real or virtual worlds. This broad area of enquiry meant that the field was lacking a unified approach to what might explain the UVE and also lacking a systematic method for collecting empirical data about the UVE. Research into the UVE then developed some of these initial ideas by looking at evolutionary theories, neuropsychological techniques and sensitivity to detecting flaws in

appearance. This review will divide that body of literature into two sections: work which applies theoretical perspectives to make suggestions about the nature of the UVE and those which empirically tested the effect itself.

1.2: Proposed Explanations for the UVE

The idea of the UVE touches on a broad range of questions. As a product of changes in human-likeness, it suggests philosophical questions from the basic level of what it means to be human, through to primal fears of death and ‘otherness’ to concerns about emerging technologies and a changing world. As a negative emotional response it touches on a broad body of existing literature on how we experience negative emotions and how these are elicited, and how subtle emotions can be defined and measured rather than be taken for granted. However, as a psychological construct, it should be possible to define the nature of the UV response and identify stimuli that are reliably able to evoke it under similar circumstances. To this end, this section of Chapter 1 looks across research which has proposed explanations for the UVE and reviews the extent to which they can be seen as potentially plausible and good candidates for empirical exploration. The review will then move on to look at studies which have carried out empirical tests of the UVE, some of which have been grounded in these explanations.

1.2.1 Evolutionary Aesthetics

One of the most straightforward explanations for the UVE is that certain entities appear unsettling because they are not attractive from an evolutionary perspective: MacDorman et al (2009) used the term ‘evolutionary aesthetics’. Certain facial appearance traits have been found to be generally and universally attractive, for example Langlois et al (2000) found consistent ratings of attractiveness in a meta-analysis of face perception studies. Perception of attractiveness appears to be an automatic and innate

process with high attractiveness linked to reproductive fitness so being able to identify and choose an attractive mate is an adaptive ability that increases the likelihood of producing viable offspring (Little et al, 2002). There is therefore an evolutionary pressure to be sensitive to attractiveness, and to be skilled at making accurate judgments of attractiveness. By this explanation, the NHEs that fall into the UV are unsettling because they are human enough to be judged in terms of attractiveness but are not conventionally attractive, so we find them unpleasant due to these signals that they may be unfit candidates for mate selection. Schneider et al (2007) collected ratings of attractiveness and human-likeness and found evidence for an UV-type effect, in that attractiveness initially increased with human-likeness but a valley occurred where the least attractive characters were those rated between 3 and 3.5 on a human-likeness scale of 1-5.

An evolutionary explanation could also relate the UV to the concept of pathogen avoidance, building on Rozin's (1987) work on disgust, as there is an evolutionary advantage to being able to avoid contact with something that may cause disease or injury. However, Moosa et al (2010) argue that the picture is more complex, and suggest that a general danger avoidance mechanism is at work in causing the UV, rather than a specific need to avoid particular pathogens. They note that many non-human species have methods of separating the dead and the living and that several do practice burial, suggesting this is not a specifically human response. Both theories seem intuitively appealing, as there certainly seems to be something about the appearance of uncanny entities which makes them unsettling. However, the evolutionary aesthetics argument is not sufficient as an explanation of the UVE, especially as eeriness is a quite subtle response compared to terror or disgust. It may be unpleasant to feel unsettled or 'creeped out' but many people may tolerate this discomfort reasonably easily. It is argued

that if the core of the UVE is an evolutionary protective function then a more robust sensation of outright disgust or fear, leading to the rejection of the entity, should be evoked by NHEs.

1.2.2 Category Boundaries and Categorical Perception

The concept of the UVE as a result of category boundary conflict or category confusion appears intuitively appealing. Historically it has generally been the case that human beings have encountered things that are either definitely human or definitely not human so the introduction of NHEs that possess many of the qualities of a human along with non-human attributes introduces an ‘other’ category which, it could be argued, we are ill-prepared to deal with. Uncertainty over the nature of things was at the heart of early writings on the uncanny (Jentsch, 1906) and since then, this concept has received considerable attention from researchers. Ramey’s (2005) work has been described in in Section 1:1 but to recap the key points, he suggested that a UV could occur for any case where there may be confusion about the point at which something becomes the member of one category as opposed to another, similar, but different category. He argued that there are many examples where this might occur but questioned whether there was anything special about the example of human-likeness except for the fact that we are particularly familiar with humans and therefore may be more sensitive to the qualities that distinguish the human from the non-human. The idea of boundary distinction causing an unpleasant sensation has also been used to explain the UV by applying the theory of cognitive dissonance to the effect. Plantec (2007) does not cite any experimental evidence for this link but the example of an uncanny robot having both human and non-human qualities is explained as a result of the dissonance between the two categories. Kang (2009) interpreted the uncanny response to NHAs as a category-

boundary challenge that causes an unsettling effect because it threatens our sense of safety and comfort. He describes a grid of potential responses to NHAs and notes that the uncanny responses seem to arise in conditions where there is a novel experience (seeing a new type of entity which has both human and non-human qualities) but only where the viewer lacks control over the outcome of the encounter and therefore feels threatened. In a setting where the viewer has control over the outcome, horror and eeriness can be replaced by fascination and wonder. In this interpretation, the effect of the category confusion is mediated by other aspects of the setting, suggesting that the uncanny experience does not have a direct cause but may arise through an interaction between circumstances and potential outcomes. As will be seen in Section 1.3 below, recent research into the UVE has taken the concept of category boundary distinction and used this to good effect in pinpointing transition points between artificial and human.

1.2.3 Neuropsychological Perspectives

The following neuropsychological explanations suppose that there are automatic, low level processes that respond differently to NHEs compared to humans or non-human entities. The following three studies offer potential neuropsychological explanations for the UVE. Hanson (2003) proposed a neural explanation for the UV which combined known neural mechanisms involved in processing emotional expressions with a proposal that near-human faces (NHF) disrupt this by not providing all the necessary cues for social engagement. He described the differences between the processing of fearful and happy faces whereby a fearful expression would trigger an alarm response in the observer, readying them to respond to whatever was causing fear in the person we are observing. The happy expression would cause no such alarm and the observer would instead prepare to engage in some form of social behaviour. However, NHEs would begin

to activate those social engagement cues because of their human-like appearance but then an alarm response would be triggered because some non-human aspect of their appearance would contradict or not fulfil the normal cues required for social engagement. This theory, termed the 'verisimilitude discrimination response' (VDR) proposed that the UVE is the result of a distributed network of various brain systems reacting to partially-satisfied expectations. It was mainly synthesised from evidence drawn from studies of activity in the fusiform face area and amygdala in response to different facial expressions and particularly in response to crisis stimuli. The conclusion of Hanson's paper was to recommend detailed imaging studies to validate the VDR theory and several preparatory experiments were also proposed to provide baseline evidence of the VDR from standard psychological experiments, and to provide preliminary evidence that would guide the eventual imaging studies. While the proposal is certainly intriguing, that lack of any experimental evidence to support whether the VDR could indeed be an explanation for the UV means it is not possible to evaluate the validity of this theory.

Lateral inhibition was explored by Shimada et al (2007) who suggested that the inhibition caused by seeing an almost-human robot could cause any pattern representing a human to be suppressed. This would cause the robot to be recognised as an object, rather than as a person. It is this mis-categorisation due to the activity at a neural level that may cause the uncanny sensation. Shimada et al presented the results of two behavioural studies: in the first study, participants held a conversation either with a human or with one of three humanoid robots varying in their human-like appearance. Participant gaze response was measured and the results indicated that gaze responses were the same for the near-human robot and the human interlocutor, but markedly different for the two less human robots. Their second study measured preferential looking towards the different androids in infants aged 12 months, 18 months and 24

months. They interpreted the pattern of gaze responses as representing another type of UV where the youngest participants were unconcerned about the android, the 18 month olds alarmed and the oldest children were again unconcerned by its appearance. This age-related difference is linked to the development of lateral inhibition mechanisms and suggests that this is the source of the UVE. However, behavioural studies do not directly measure lateral inhibition, they only extrapolate that it may be occurring differently at different levels of human-likeness.

Saygin et al (2010, 2012) studied the action perception system (APS) in relation to android perception. The APS perceives body movement and has been localised to a network of lateral superior temporal, inferior parietal and inferior frontal brain areas. These include mirror neurons (Keysers & Gazzola, 2005) which respond or resonate when people see an action being performed, as well as when they are performing it directly. Their study measured responses to video clips of an android which could appear either mechanical or highly humanised. Two sets of video clips of the android performing everyday actions were made, allowing comparisons to be drawn between the mechanised or humanised version as well as the original human model on which the android was based. fMRI measurements of repetition suppression were taken. It was found that there was a difference in the level of suppression found for the three levels of realism: repetition suppression was found in the posterior temporal cortex for all videos but there was a larger and more distributed pattern of suppression for the realistic android compared to the human or the mechanical android. They also found that there was considerably less activity in the left hemisphere when viewing the videos of the mechanical, robotic-appearing android. Saygin et al noted that this could suggest lower activity in the Extrastriate Body Area which responds to perceptions of the human body as the mechanical android lacked a human appearance even though it was making

human-type movements. They described the pattern of activity as representing uncanny 'hills' - unlike the drop in emotional response characteristic of the valley. Their findings represented an increase in activity for the human-like agent compared to the real or mechanical version of the agent. The authors noted that they did not explore the emotional component of how participants felt when they were viewing each of the videos so this study cannot comment on the eeriness aspect of the valley, but it does provide evidence of that brain activity that varies according to the realism of the agent being witnessed. There is some evidence from these three studies that there are particular structures involved in the perception of human-like entities but these explanations do not constitute evidence of a neural basis for the UVE, nor are they attempting to provide an explanation of how the effect occurs.

1.2.4 Fears of Own Mortality

The social psychological Terror Management Theory (TMT) was used by MacDorman (2005c) to link the unsettling components of the UVE to associations with personal mortality. TMT proposes that being aware of one's inevitable demise is so distressing that psychological defences need to exist to allow people to deal with this fear so it does not become overwhelming. The theory suggests that conscious thoughts are dealt with by rationalization or suppression. These defences involve the development of a view of the world which provides and promotes rationality, continuity and organization, and acts as a barrier between oneself and the inevitability of death. The 'mortality salience' hypothesis is used to test this idea as anything that triggers fear of death should also increase support for that worldview. This is presented as a 'distal', non-conscious process where even subliminal death-triggering cues should elicit a tendency to prefer that which is

culturally familiar and comforting and to avoid anything that might threaten this view of the world.

MacDorman (2005c) used an online experiment to test whether these distal mechanisms of worldview acceptance would differ when participants saw a human-like android or a female actor. The humanoid android had superficially similar features to the real female, but also had a slack mouth, blank eyes and disconcerting 'gaps' around the eye sockets. All measures were based on earlier research on mortality salience which had found that thoughts of death triggered a preference for charismatic leaders over relationship-oriented leaders, a preference for praise for one's country over criticism, and a tendency towards negative interpretations of words which shared a common stem but could be completed either negatively or positively. For example, e.g. MUR--- could either be completed 'murder' or 'murmur', --EOPY could either be 'creepy' or 'sleepy'. The results showed that the participants who saw the android rated the charismatic ('worldview-supporting') speakers significantly better than the relationship-oriented speakers, and the praise over criticism. They also chose significantly more of the death or uncanny spellings for the word puzzles compared to those who saw the real human image. These findings are what would be expected by the mortality salience hypothesis if the android was triggering thoughts of mortality, and MacDorman et al concluded that these results may demonstrate that the eeriness of seeing an uncanny android does result from a fear of death.

The usefulness of this study does rely on an acceptance of TMT, and that is an area in debate. Navarrete and Fessler (2005) argued that the foundations of the theory itself are dubious, and that any evolutionary mechanism requiring such complex defence mechanisms is unlikely to ever be a reasonable adaptation. In addition, MacDorman et al acknowledged that these findings were based on only one example of an uncanny

stimulus, and suggested exploration should take place across a wider range of ‘uncanny’ stimuli. The issue of identifying what constitutes an uncanny stimulus will be examined in detail in Chapter 2: Methodology.

Four theoretical perspectives have been considered as ways to examine the UVE. However, they either did not offer empirical support for their proposals, or offered support based on observations from examinations of individual NHAs. This review will now look at empirical studies which have been carried out to test the concept of the UVE across different approaches.

1.3: Empirical Tests of the UVE

The following section considers seven areas of research into the UVE by evaluating the empirical evidence that has been presented for the presence of an effect and its cause. Four of these areas have been introduced in Section 1.2 above and concerned evidence that the UVE results from an evolutionary pressure, ambiguity at category boundaries, from neuropsychological structures and processes or from a fear of death. Four new areas are introduced below: empathy and trust; error sensitivity; anomalous features; and cue mismatch effects. A characteristic of studies conducted within these areas is that they all use more than a single example of human-likeness to explore the UVE as it might apply more generally, rather than in response to just a single stimuli.

1.3.1 Evidence of Evolutionary Aesthetics

Section 1.2.1 described evolutionary theories that had been proposed as explanations for the UVE. Steckenfinger and Ghazanfar (2009) was the first study to use non-human subjects to validate an evolutionary explanation and they argued that if the valley emerged in non-human subjects, it could not be said to rely on specifically human processes and therefore the evolutionary explanation may be valid. This study consisted

of preferential-looking behaviour in macaque monkeys who looked at images of monkey faces that were either CG or photographs. The CG faces were rendered at two levels of detail, one highly realistic and one unrealistic. Steckenfinger and Ghazanfar hypothesised that spending a shorter time looking at the realistic faces compared to the unrealistic or real faces would indicate a UV pattern of responses. The summary of the potential outcomes is shown in Figure 2 below and is illustrated as a useful way to visualise the potential outcomes to the different stimuli:

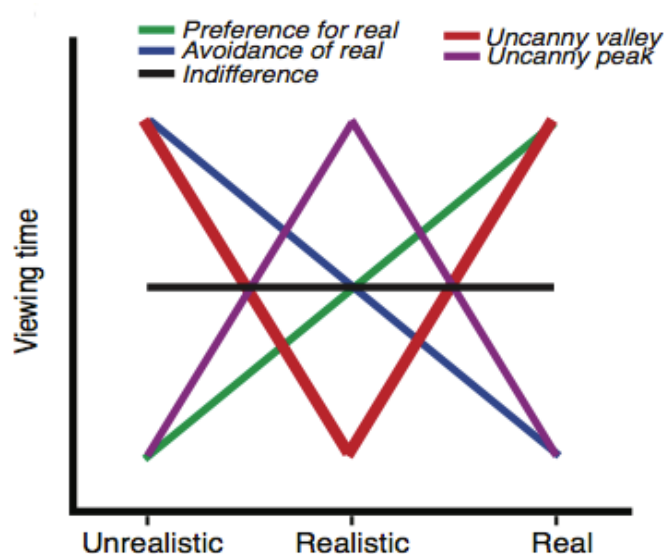


Figure 2: Potential outcomes from a preferential looking experiment with three levels of realism. (From Steckenfinger and Ghazanfar (2009)).

Their results supported the UV outcome as the monkeys spent significantly less time looking at the realistic faces compared to the unrealistic and real faces, measured by duration of looking and by number of fixations. This pattern occurred whether the faces were static or dynamic, when different facial expressions were used in the stimulus images and when the images were presented at different levels of screen brightness. Interestingly, and counter to Mori's original theory, there did not seem to be a significantly deeper valley for the moving faces compared to the static faces. This pattern of results supports the notion that the UV is not uniquely a human phenomenon and

begins to support the evolutionary argument for a UV. However, as the authors note, it is not possible to tell why this response pattern occurred as data which could suggest whether the subjects were alarmed or disgusted by the realistic faces were not collected as part of this exploration. The study indicated that a response pattern suggestive of the UV is found more broadly than in humans, but not whether there was necessarily any emotive component to the reason why the monkeys preferred to look at the non-realistic and real faces. It could be argued that they might have found those faces more attractive so looked at them more, but the non-realistic faces were rather alarming in appearance which makes this hard to support. However, it is certainly a valuable initial indication that supports the idea of a valley type effect in non-human animals and that some aspects of the evolutionary arguments may indeed be sound.

Following on from this evolutionary evidence, Lewkowicz and Ghazanfar (2011) argued that the common pattern of responses between monkeys and humans may be because developing the ability to process and recognise faces requires repeated exposure to face stimuli. At some point during the development of this ability, infants move from being drawn to general representations of face-like stimuli to a specific preference for faces that have particular relevance to their situation and culture. Lewkowicz and Ghazanfar (2011) suggested that it is during this perceptual narrowing that the UV emerges and carried out a study of infants aged 6, 8, 10 and 12 months old. They suggested that the UV should emerge between 10 and 12 months of age. To explore this, they created pairs of faces at three levels of human-likeness: a human face, a realistic avatar and an uncanny avatar which was based on the realistic avatar but had been modified to give it unsettlingly large eyes. It was found that 6-month-old infants preferred to look at the uncanny face over a human face but the human face was significantly preferred by 12 months olds. Looking preferences had crossed over in the second six

months of life, suggesting support for the theory that perception of the UV develops rather than is an innate perceptual quality.

It can be concluded that there may be an evolutionary and a developmental component to the UV. In terms of response pattern at least, non-humans experience it and it is not experienced by very young children, but emerges during the second six months of life. However, these studies did not examine the emotional aspect of the UV response, nor did they suggest any mechanisms by which the response might be explained.

1.3.2 Evidence of Category Boundaries and Categorical Perception

Section 1.2.2 above suggested that one explanation for the UVE may be difficulties in making categorical decisions about the nature of the NHAs. Matsuda (2012) et al considered the question posed by Ramey (2005) about the issue of valleys for categorically distinct entities other than humans vs non-humans, and asked whether a valley would occur between categories of familiarity and novelty in infants' preferences for looking at images of their mothers or strangers' faces, compared to CG images created by morphing the mother/stranger images together. Three images from the morph sequence were used: the initial mother image; the final stranger image; and the intermediate morph between the two. Three age groups (7-8 months, 9-10 months and 11-12 months) were involved in the study and it was found that a UV-type effect did appear irrespective of age. This manifested as the infants all preferred looking at the mother and the stranger more than the intermediate image created by morphing the two together. The mother was slightly preferred over the stranger. However, it could have been that the morph itself was off-putting for the infants, causing them not to look at it for so long. To test whether the results in the first study did occur as a result of the valley

between familiarity and novelty, a second study was conducted where Matsuda et al also morphed images of different strangers together and presented the resulting morphs of those to the infants. The results of the second study showed that there was little difference in looking preference between the morphs and original stranger images, meaning that the valley effect observed in study one did not seem to occur. They concluded that a valley does exist between familiarity and novelty, but as with the non-human studies presented above, it would not be possible to use this evidence to look at the emotional component of that response.

Cheetham et al (2011) applied the concept of categorical perception to the human-likeness axis of the classic UV graph, questioning whether the steady and smooth continuum of a dimension of human-likeness as presented on the graph was actually psychologically plausible. If categorical perception is a feature of human-likeness, pairs of images that 'straddle' a category boundary should be easier to distinguish compared to pairs taken from within either category. Their results supported the hypothesis that the measures of human-likeness for morphed faces are categorically perceived, as pairs between the categories were significantly more likely to be judged as different than pairs within the categories. They concluded that their findings could be used as a framework for a psychological exploration of the UV and may offer a way of locating Mori's theory within a tradition with existing parameters to help direct future research.

Cheetham et al (2013) went on to explore the 'dimension of human-likeness' using eye-tracking techniques to investigate whether a relationship would be found between fixation and categorical perception. They suggested that a longer dwell time on faces near to the categorical boundary would support the additional cognitive load created by making a categorical decision under uncertainty. Cheetham et al (2013) also looked at the number of fixations for different facial features. While they did not find significant

differences in the number of fixations on faces at the category boundary compared to faces clearly within a category, they did find that a feature hierarchy emerged with more fixations on the eyes than on other features. In terms of dwell time, participants did spend significantly longer looking at faces at the category boundary, providing support for the hypothesis that processing faces at this particular boundary requires greater perceptual effort to extract the relevant information in order to make the distinction.

The evidence for categorical perception as a factor in the valley-type response to agents varying in human-likeness has been considered and it seems clear that ‘something’ occurs at the liminal boundary between human and artificial and framing this in terms of categorical perception theory may help to explain the valley effect, but not why it is particularly uncanny in nature. As Cheetham et al (2011) noted, the category perception theory of the UV does not attempt to offer a commentary on the emotional component of the effect and there is not yet any explanation for why the increased sensitivity at this boundary should produce an uncanny sensation. Therefore, there are definitely more questions to ask about the qualities of entities on the boundary between human and non-human and the effect that they can have on how people feel.

1.3.3 Neuropsychological Evidence

Section 1.2.3 above presented theories for the neuropsychological basis of the UVE, and some studies have collected measurements of neural activity at different levels of human-likeness which may help to develop an understanding of the effect of changing human-likeness. As described previously, Chaminade et al (2007) compared responses to animated short films of characters which varied in the extent to which they had realistic human characteristics, starting with simple shapes composed of individual ‘dot’ point-light displays or jointed ellipses, to a mechanistic-looking robot, through to a humanoid

monster and a human-like clown and jogger. The characters were either created from motion capture, representing natural motion, or CG, representing an artificial animation. Participants viewed a random presentation of the characters and were asked to decide whether they were seeing the natural motion version or the artificial animated version. Bias towards reporting natural motion and fMRI measurements were taken. A difference in this bias between the different types of characters was found with the dot displays showing the greatest increase in 'natural' bias and the robot and clown/monster representations showing the least. This demonstrates that the movements which were presented with the human-like characters were judged as less natural than those with only rudimentary human characteristics. Chaminade et al linked this increasing negative bias to the idea of the UV, as the animated characters did only provide a partial representation of human-likeness as the only animated aspect was the running motion, so while that movement would have had an element of realism it could have contrasted awkwardly with the still face and blank expression.

The imaging results were used to examine this idea, and it was found that the negative response bias correlated with a decrease in activity in regions belonging to the mirror system, and an increase in regions involved in mentalising. They gave one explanation for the decreased mirror system activity as being due to the lower human-likeness of the stimuli and that a more realistic depiction of a human may be required to trigger the activity of mirror neurons. The increase in mentalising could be due to a variety of factors, but they suggested that a lack of familiarity with the stimulus might be the cause. This study supports the idea that there is a non-linear relationship between human-likeness and reaction, and does provide some support for the type of relationship Mori proposed. However, no emotional measures were made of how participants felt

towards the different levels of realism so while an effect is present, its nature as uncanny is still unproven.

1.3.4 Trust, Social Actions and Empathy

A concern arising from the development of social robots is whether their users or owners would be able to interact with them comfortably or whether aspects of their appearance would make them too unsettling to interact with as a social entity. This concept has also been explored in the use of artificial agents in computer-mediated interactions. The following section will describe research into the role of trust and social actions in NHAs and will explore the implications for the UVE.

Mathur and Reichling (2009) explored social response as a function of human-likeness by presenting participants with images of robot faces which varied from mechanical to human-like and realistic. As an interesting methodological note, they considered the use of morphing between human and robot faces to create their continuum but as their work was grounded in a robotics tradition and their intention was to be able to make design recommendations for 'buildable' robots, the decision was taken not to use this technique. Instead, images were collected of robot faces that varied in the degree to which they displayed human-like characteristics. They measured social attitudes to the robots in two ways; firstly, by using a survey to collect rating information on whether people would like to interact with the different robots and then by a modified game theory methodology where the trustworthiness of each robot could be ascertained more directly. The results of the rating survey found a pattern consistent with the UV theory, with an initial rise in acceptability followed by a decrease below neutrality to negative liking when the images became more human-like, with a final rise again as they became human in appearance. In the game theory experiment, they found that the near-human robots were the ones likely

to receive the lowest bet and that this could be significantly influenced by small modifications to the facial expressions displayed by the robots. They concluded that there does appear to be a UV for social attractiveness and that this represents a real-world concern for designers working to produce robots that will need to interact with humans.

McDonnell (2010) measured how the trustworthiness of virtual agents when they were presented at different levels of realism: a high-quality, near-photographic quality render; a lower quality render corresponding to that used in most video games; and a non-photorealistic render which appeared similar to a ray-traced caricature of a face. Participants were significantly more likely to conclude that the faces rendered in the highest quality (i.e. almost but not quite human) were lying. This suggests a lack of trust for NHAs compared to ones rendered at lower levels of realism. It would have been interesting to compare how trustworthy participants found the actual human model who had been used to record the statements and after whom the renders had been modelled but this comparison was not included in the experiment. This lack of a natural human anchor for comparison of any NHAs will become an important methodological consideration that is explored later.

In anthropomorphising or attributing human qualities to a non-human entity, one is making an assumption that the entity shares something in common with the observer and as such a sense of misplaced empathy occurs where the entity seems human but is not. To what extent can our tendency to attribute human-like qualities to non-human and almost-human agents explain the unsettling qualities of the UVE, and is this tendency affected by our perceptions of animacy and mimicry?

Woods (2006) was interested in the range of robot designs that children would find acceptable, both as desirable entities to interact with in play but also to explore whether a UV might make some robots off-putting. Forty images of robots were categorised as

human-machine, animal-machine, and machine-like and rated using a questionnaire which included items about the robot's appearance, personality and its emotions. A principal component analysis of the correlations between the mean scores produced two factors which were labelled 'behavioural intention' (BI) and 'emotional expression' (EE). BI was linked to perceptions of the robot's intent, either positive or negative, and EE was linked to the perceived happiness or sadness of the robot. Exploration of the two factors did show tentative support for the UV, as ratings indicated that the children were least comfortable with robots which were human-like but not 100% human. Those robots that were categorised as 'human-machine' in appearance had the highest positive BI scores, but those that were rated as 'human-like' had the lowest. Woods noted that there had been no deliberate measurement of how close to human each of the robots was, which made the positioning of each example of the horizontal axis of the UV graph problematic, and recommended that this should be investigated in detail in a separate study. It was concluded that NHAs are subject to anthropomorphism and are perceived as happy, sad or angry based on the non-emotive components of their appearance.

Riek and Robinson (2008) examined empathy with NHAs by adapting a commercially available robot to be capable of facial expression mimicry. The authors described how being able to mirror the emotions displayed by someone else in social interaction aids with the establishment of empathy because mirroring someone's expression indicates non-verbally that you understand what they are saying and to an extent are at least capable of experiencing the same emotions even if they are not being elicited at that moment in time. Riek and Robinson found that participants reported most satisfaction with an encounter with the robot when it attempted to mimic their own expressions. Participants were also interviewed and those findings indicated that the mimicry condition increased empathy with the robot. In terms of the UVE, this suggests that an

absence of mimicry may be one of the factors involved in causing NHEs to be disliked. If the agent cannot convey that sense of empathy, the interaction with them would not seem natural and may lead to feelings of unease.

However, there may be a limit to this. Leander et al (2012) found that excessive amounts of mimicry could actually prove negative rather than positive. The researchers observed that implicit behavioural rules tend to be well followed during normal interactions, but any violations of those standards could cause a negative reaction in an interlocutor. This negative response can be characterised by experiencing feelings of physical coldness towards the person violating the standards of behaviour. This is described figuratively when someone whose behaviour is 'a little off' is said to give someone 'chills' or 'sends me cold'. The researchers explored this experimentally in studies where mimicry was either absent or displayed inappropriately and found that there was a complex pattern for where mimicry appeared to ease the social interaction and when it did not. It was not the case that mimicry was always welcome or unwelcome, but instead it varied according to context. For example, when encountering a friendly interlocutor, a lack of mimicry caused feelings of coldness but when the interlocutor was task oriented and not as friendly, mimicry caused feelings of coldness. The picture was complicated further by the finding that there were no absolute thresholds for acceptable levels of mimicry, but that these actually varied from person to person. Applying these findings to the UV, it was suggested that it would be very difficult for a NHA to interact in a way that did not make people feel chilled and cold as this would need to be extremely sophisticated at reading and responding to the human cues to be able to evaluate and appropriately return mimicry behaviours that were timed appropriately, judged correctly for the situation and were appropriate to that particularly individual.

Gray and Wegner (2012) suggested that human-like robots may be unsettling, not because of some isolated aspect of their appearance, but because they give the overall impression that they might possess a mind and it is this impression that sets up expectations which then cannot be fulfilled. People perceive the presence of a mind in another as either a dimension of agency, where the other has the ability to plan and act, or as a dimension of experience where the other is able to engage with the products of its actions and experience the world. Gray and Wegner used a series of encounters with a robot which could appear very mechanical or very human-like, where the amount of agency and/or experience could be varied. Measurements were taken of the extent to which the robot was judged as uncanny. The researchers found that when participants saw videos of the human-like version of the robot, they reported that they were more unsettled, uneasy and 'creeped out' than when seeing the mechanical version. No differences in the agency perception were found between the two levels of human-likeness, but the human-like robot was significantly more likely to receive attributions of experience compared to the mechanical robot. A second study tested whether we apply such attributions to humans as well as robots by presenting subjects with a description of a person who had lost abilities relating to agency or experience or was not affected by lost abilities. Participants provided ratings of eeriness for the person under each of the three conditions and it was found that the human lacking in experience made participants significantly more uneasy than the human without agency. In other words it is more eerie to encounter a person who cannot feel than one who cannot act. The researchers concluded that the entities that fall into the UV do so because they are human-like agents who lack an ability to experience. As long as robots are mechanical, their appearance remains utilitarian and therefore we are comfortable with them as 'things that do', but

when their appearance becomes human-like and suggests they should also be able to experience, this is no longer acceptable and causes unease.

However, an aspect of empathic behaviour which was not found to be affected by levels of human-likeness was the ability to take on another's viewing perspective. MacDorman et al (2013) described a study looking at the ability to take on the perspective of someone (or something) else. A prerequisite of this ability is being able to imagine a scene from the point of view of an observer placed somewhere in the scene rather than how it appears to the participant themselves. This 'perspective taking' has been used as a measure of empathy as it requires the establishment of a theory of mind in the case of the observer. The researchers hypothesised that there would be a difference in the extent to which participants were able to take the perspective of observers of different levels of human-likeness and that levels of human-likeness close to the UV would cause difficulties in the ability of participants to take their perspective. This hypothesis would seem to follow from the observations described so far but the final results showed that there was no such difference as uncanniness did not influence perspective taking ability. This does not invalidate the areas of empathy where human-likeness did have an effect but does suggest one area where it appears there is no relationship with human-likeness.

One final aspect of mind perception that has been considered in this area is that of animacy. Perceptions of animacy can be seen as a necessary condition for empathy to occur as it is only at the point where we perceive something as alive that it becomes a candidate for interaction. Looser and Wheatley (2010) explored this experimentally in a series of studies where morphed images of faces were presented to participants who were asked to identify the point at which they perceived the face as animate. There were three key findings from the studies. Firstly, perceptions of animacy and a sense of mind

consistently appeared towards the human end of the continuum. They looked for a point of subjective equality (PSE) where participants judged that an image representing a specific point in the morph sequence was equally likely to seem animate as inanimate. As the morphs were created by transforming inanimate faces into animate ones, this point should appear at the 50-50 point between the two. They found that it actually occurred significantly closer to the human end of the morph, between the 60 and 65% human points. Secondly, they found that participants (who had not rated the original images for animacy) were able to categorise pairs of images that straddled the PSE as animate or inanimate, suggesting that their appearance contained information that allowed participants to put them into qualitatively different categories. Finally, when the face images were processed to present just part of the face, the eyes were found to be most salient when participants were making judgements of animacy. These findings are applicable to the UV idea as they found that on a continuum from inanimate to human, animacy emerges just before the uncanny dip is thought to take place. They did not specifically test for eeriness but their discussion includes an anecdotal report that some of the participants found the morph faces at the animacy boundary to be unsettling. With perceptions of animacy appearing just before the hypothesised location of the UV, it may be that a perception of animacy is required for us to begin to perceive faces as close to humanlike and thus for uncanniness to occur.

While the role of empathy and the tendency to empathise with others may be a factor in understanding unsettling responses to NHAs, the research reviewed here demonstrates an incomplete picture. It is suggested that we do not trust almost-human agents or find them credible and there is a definite point where an inanimate object begins to appear animate. However, some effects which may be expected with a NHA do not manifest. For

example, mimicry creates empathy but only up to a point and there is no ‘valley’ apparent in perspective taking during interactions with a NHA.

The studies discussed above suggest that the UVE may result from a difficulty in trusting agents that appear almost but not quite human. The question of how comfortable people would feel with a NHA was used to frame one of the research questions for the current project in which a scenario of how comfortable participants would feel about sharing their home with a social robot of different appearances was used to elicit emotional reactions to NHAs.

1.3.5 Error Sensitivity

It has been suggested that the UVE occurs as a result of increased realism causing greater sensitivity to the non-human elements of an agent. In other words, as NHAs become more realistic there is more information available from their appearance and behaviour which means we may be more sensitive to any small deviations from genuine humanness that designate near-humans as different, dangerous or alarming. These deviations therefore acquire a salience out of proportion to the size of their defects and become the focus of attention (Steckenfinger and Ghazanfar, 2009). MacDorman et al (2009) found evidence for the role of error sensitivity when examining the level of detail used to display a face, and found that two subtle manipulations of the facial proportions (eye height and eye separation) would interact to cause significant differences in whether the faces were judged as eerie. The level of detail was operationalised by three types of texture being used on the faces: line drawings; a bronze texture; and a photorealistic rendering. The study was carried out online and the researchers employed a novel method for operationalising eeriness, in that rather than rating the images on a scale, participants were presented with an interface where they could modify the level of detail

and were instructed to do so until the faces were as eerie as they could make them. They found that sensitivity to the eeriest proportions was indeed highest when the level of detail was highest, suggesting that the idea of error sensitivity is a valid one. However, it should be noted that only three levels of realism were used in this experiment and the base model (a static male figure with a neutral expression) was the same for each pose so it would be difficult to generalise from these observations. Tinwell et al ((2011a),(2009)) also identified error sensitivity as a salient factor in causing the UVE. Participants rated videos of virtual characters on measures of familiarity and human-likeness and it was found that rather than describing a single valley-type curve, a series of local minima were found when familiarity was plotted against human-likeness. They suggested that rather than a UV, an 'uncanny wall' would be a more apt description for the response pattern observed in their studies and theorised that it may never be possible to traverse the UV as people would begin to adapt to each new development in realism and that the sensitivity to the anomalies in the artificial agents would always be marked and significant in making them seem unsettling and eerie. There are positions that run counter to the error sensitivity argument and this includes Pollick (2009), who asked whether it really is inevitable that increased realism would lead to an increase in sensitivity to errors, and questions the mechanism by which this might occur. It seems just as plausible that the more realistic figure may present so many cues that are realistic that they 'drown out' the anomalous information rather than emphasise it. In light of this counter argument, error sensitivity does not seem an explanation in its own right and it may instead function as way to describe the uncanny problems, as it does serve to present some elements of the effect effectively.

1.3.6 Anomalous Features

The explanations and evidence considered to date have mainly been concerned with holistic traits of NHAs which might be responsible for their unsettling effects. NHAs disturb us because they may call to mind thoughts of our own mortality and make us feel threatened, they make us believe they are inanimate when they are not or they cause consternation because they are difficult to assign to an ontological category. These general explanations tend to consider the agent as a whole entity but the last two sections of this part of the review will narrow the focus to the appearance of the NHA itself and consider its component parts to explore whether the uncanny experience can be said to result from features that are anomalous in appearance or unusual in their proportions. This can be observed anecdotally when presenting potentially uncanny stimuli to peers as people often responded with exclamations of alarm at the appearance of the eyes, which were then described as cold, blank, unsettling or even 'evil'. However, there are more rigorous sources of conjecture on this issue. Brenton et al (2005) suggested that an explanation for the eeriness of the UVE was that the appearance of NHAs conveyed impressions of falsehood or deceit. An artificial agent's inability to convey a fully-realised and genuine expression in a context where other aspects of its appearance suggest a believable character may be off-putting and unsettling in an interaction. They argue that the root of this effect would be present in the eyes of faces given their salience for communicating intent and emotional expression. Fear may be considered a sensible response to an inability to read the intentions of another being.

Geller (2008) addressed the importance of realistic eyes in NHAs. In a section focusing on the animation techniques that can be used to create a believable representation of a human character, he quoted several animation practitioners on the challenge of reproducing the fine details of attention, intention and emotional expressions that are

carried mainly or solely in the eye region of an artificial face. Geller suggested that being able to present realistic eyes is key to avoiding the UVE in animated characters to the extent that some productions splice together footage to use videos of real eyes in a CG face to overcome the 'cold' effect that results from eyes that are not realistic enough to be acceptable.

To discover the parameters for acceptable CG characters, Green et al (2008) allowed participants to manipulate the proportions of images of faces to either create the most or least acceptable configuration of individual features. The faces were presented at different levels of human-likeness so the acceptable proportions for very human-like characters could be compared to those for mechanical robots and other far-human agents. They found that the more human-like the faces, the narrower the range of tolerable proportions with only small changes being required on the human-like faces to make them lose acceptability. In terms of uncanny agents, this supports the concept that their eeriness may be due to small deviations from acceptable proportions. It has been suggested in Section 1.3.5 above that one reason for rejecting a NHA could be that its human-like qualities do not match with the qualities that are not sufficiently human-like and this may be an explanation for that finding.

Another aspect of MacDorman et al's (2009) research, which was described above, considered whether participants would be more likely to agree on acceptable proportions for faces that were closer to human-like compared to those with lower levels of detail or which possessed features that made them look less human. Nine different CG characters, representing all possible combinations of appearance based on the different levels of texture and detail, were presented to participants. These were three styles of texture representation (a wireframe sketch, a smoothed bronze statue-like texture and a photorealistic skin-like texture) with each presented at three levels of detail from a basic

realisation through to a detailed rendering. They found that there was a difference in the acceptable proportions as the figures became more human-like. With regards to the UV theory, this supports the idea that part of the unsettling effect may be when some features do not meet those standards of acceptability.

Finally, Seyama and Nagayama (2009) developed the concept of acceptable proportions in more detail in a series of studies systematically varying eye size and examining the effect that this had on how unsettling people found the resulting faces. They carried out three experiments using an adaptation technique where the adaptation faces had been manipulated to increase the size of the eyes to disproportionate and unusual-looking levels. The adaptation paradigm anticipated that after looking at these adaptation faces, participants looking at non-manipulated test faces should report that the eyes look disproportionately small. The faces used in this study were six pairs of images. Images of six humans were matched to similar dolls or masks, and manipulated to produce three sets of stimuli consisting of natural faces with eyes increased to 150% of their normal size, artificial faces with eyes at 150% and artificial faces with their eyes hidden. These faces were split up into adaptation and test stimuli, with participants being required to adapt to one of the three sets and then tested using the natural set. The aim was to find out where the boundary of acceptability for eye size would fall for each participant after they had been adapted to the manipulated face. Would the participants who saw the enlarged eyes on the artificial faces report that smaller eyes on the natural test face were acceptable compared to those who had been adapted to the natural faces or to an artificial face with their eyes hidden? After a baseline exercise was carried out to ascertain their threshold for non-modified eyes, participants were presented with their three adaptation and test faces and their mean boundary size measured. The boundary size was identified using a PEST detection method to select the smallest acceptable

difference in eye size. From the baseline measurement and adaptation/test boundary, an eye size aftereffect could be calculated. The results of this demonstrated that there was a significant difference between the artificial, natural and artificial/eyes hidden faces, with the natural faces causing the largest aftereffect compared to a very small after effect for the artificial faces. The artificial/hidden eyes condition also caused a small aftereffect, well below the natural faces. The researchers offered two interpretations for this finding. Firstly, that there is a different mechanism involved in the way that artificial and human faces are processed so that seeing a natural face was not affected by the prior sight of the artificial face. Alternatively, it could be that this adaptation effect is transferable between artificial and human faces, but the conditions of the study were not sufficient to elicit it. This possibility was explored in their second study where the artificial faces were used as test stimuli. The eye-hidden condition was also dropped from this second study. 9 found that when tested on artificial faces, the after-effect for human faces did occur, countering the suggestion that there was a different perceptual mechanism in place for the artificial and human faces. The same pattern of results was found when they modified the experimental conditions used in the first experiment to make the task easier and the adaptation period longer. Their conclusions were that when explored using eye-size adaptation effects, human and artificial faces were processed in the same way. However, this does leave the outstanding issue of NHFs and whether the same perceptual mechanism would have been applied if rather than the artificial and human faces, an almost-human face from a morph sequence between the two had been presented.

It can be seen that explorations of the role of how we perceive different facial features and the influence of strange or anomalous features can provide valuable insights into a sense of the uncanny. To progress this area of enquiry the present project has framed a research question developing the approach taken by Seyama and Nagayama in

applying a standard face-processing methodology to entities along the continuum of varying human-likeness to explore whether there is any evidence to suggest that NHFs are processed differently from human or artificial faces. Seyama and Nagayama used adaptation to investigate eye size aftereffect, and the present research project will use the face-inversion effect (FIE).

1.3.7 Cue Mismatch Effects

So far, this review has considered a range of potential explanations for the UV, but there is still a final area to be explored. Cue mismatches may be a cause of the UVE, particularly when some aspects of a NHA convey a more convincing appearance of human-likeness than others and so the discrepancy appears unsettling or alarming.

What may be unsettling about a robot is not that its overall degree of human likeness places it in an 'uncanny valley' but that there is a mismatch among elements—some aspects of its form, motion quality, or interactivity may seem more human than others—and it is this we find disturbing. An example of this would be the human-like eyes and teeth of [a] robot combined with its absence of skin or hair and the mechanical jerkiness of its movement. (Ho et al (2008), p176.)

Within the broad concept of cue mismatch, there are several different domains of 'cue' that could be mismatched to give rise to an unsettling experience. It has been seen earlier in this review that a mismatch in realism has been implicit in the trust, empathy and social action studies where the degree of realism has been varied as a way to explore the effect on our tendency to trust or empathise with a NHA. At some point along the continuum of realism, there is some quality or trait of the agent which is no longer in keeping with the overall degree of realism and it is at this point that the UV emerges. Brenton et al (2005) suggested that eyes were particularly relevant to the UVE as they can be responsible for perceptual cues suggestive of duplicitous or misleading intent. If the appearance of a NHE is highly realistic then we might reasonably expect to be able to

infer intent from their eyes. If this does not occur, either because the eyes lack the realism of the rest of the form or because the intent is not actually there, this mismatch in expectations could be responsible for the UVE. A mismatch could also be seen as implicit in the category boundary and categorical perception theories with their emphasis on distinctiveness and discrimination. If the realism of an entity varies in a linear fashion, any difficulties in identifying the correct category may be due to a growing mismatch between the cues suggesting category A and those suggesting category B.

Seyama and Nagayama (2007) provided experimental evidence to support the effect of a mismatch between the realism of face and eyes in contributing to a sense of the uncanny. They conducted four experiments using stimuli which had been created using a morphing technique to smoothly blend an initial face into a different face. The stimuli were still images taken at even points through this blending process. In each experiment, the basic methodology remained the same: participants viewed images of faces and were asked to indicate on a five point scale how pleasant or unpleasant they found each image. First, the researchers tested the hypothesis that a dip in emotional response should emerge at some point if a face is morphed between artificial and human, increasing in realism. They described this dip as a 'negative peak'. Three sets of images were used for the first study, morphing from three artificial faces (a doll, a CG face and a cartoon) through to three photographs of matched faces. They did not find that any such negative peak occurred as increasing realism alone was not sufficient to show a UVE. However, they observed that even the 'human' anchor at the end of the realism morph was not truly realistic, given the limitations of a disembodied head shown as a low-resolution graphic.

The UVE may not have appeared because the images were only approaching the first peak on the graph and thus not getting far enough to fall into the valley. As the eyes are

the most expressive feature on the face, the second study addressed this by morphing the eyes from artificial to human separately from the rest of the face so that either the eyes were morphed and then the face was morphed, or the face was morphed followed by the eyes. In this case, a negative peak did indeed occur as in both examples, the lowest pleasantness rating was found for the greatest mismatch between artificial and human features, e.g. the 100% human face with the 0% artificial eyes, and vice versa. As the range of realism was no different to that used in the first study, the researchers concluded that the lack of a UV found was not due to the lack of the range in realism, but instead the mismatch in facial features may have been the cause thus possibly echoing perceptions of facial disfigurement or abnormality. However, if there is a link between a UV and perceptions of disfigured or distorted faces, other abnormalities should also trigger it. Their third study tested this by manipulating eye size from normal to abnormally large, while also continuing the artificial to real morphs as for the first and second studies to control for changes in realism and for mismatch. Again, a negative peak occurred, this time at the point where the face was 100% human, but the eyes were enlarged to 150%. Thus the UV seemed to be emerging with highly realistic faces with the highest degree of eye size distortion.

To explore whether there might be other reasons for this negative peak, the researchers used a separate set of rating experiments to test for significant differences in the perception of realism, gender, expression and age, and only found a significant result for the perception of age. This led to their final study where the stimuli were deliberately drawn from a range of ages and expressions across both genders. The morphs again included both realism and eye size, which allowed the influence from the confounding factors to be subtracted. The results showed that even where the confounding factors were controlled for, there was still a negative peak for the highly realistic faces with the

most highly distorted eyes. The overall conclusion drawn from the four studies is that a UV can occur, but the sense of uncanniness is enhanced under the specific circumstances where there is an element of distortion and mismatch along with the transformation from artificial to human-like.

This finding was developed by MacDorman et al (2009) who considered the same type of mismatch, namely that between the realism of the face and eyes, could be a causal factor in determining the eeriness of CG characters. The realism of the eyes and the surrounding face of the character were modified separately to produce stimuli where the face could be highly detailed yet the eyes could be presented at a minimum level of detail. Their hypothesis was that a mismatch in the level of realism would cause an unsettling sensation and this was found in their results of analysing semantic differential scales measuring eeriness, artificiality and attractiveness. It was found that the extremes of mismatch did result in negative evaluations on all scales.

A large body of work which commenced with the consideration of a face and voice mismatch has produced some compelling evidence to suggest that this can also be responsible for the emergence of an uncanny sensation. Tinwell has explored this area with findings related to speech synchronisation, speech speed, facial appearance qualities and emotional expressions produced. This review will focus particularly on the mismatches in experience and expression which began when Tinwell (2009) looked to recreate Mori's (1970) UVE curve by collecting ratings of satisfaction, human-likeness and familiarity for virtual characters. There was no single UV curve demonstrated by plotting human-likeness against familiarity, and instead a pattern described as a 'series of valleys' appeared where there are several dips in familiarity at different points of human-likeness, seemingly clustered by groups of characters. Three distinct valleys appeared as can be seen in Figure 3.

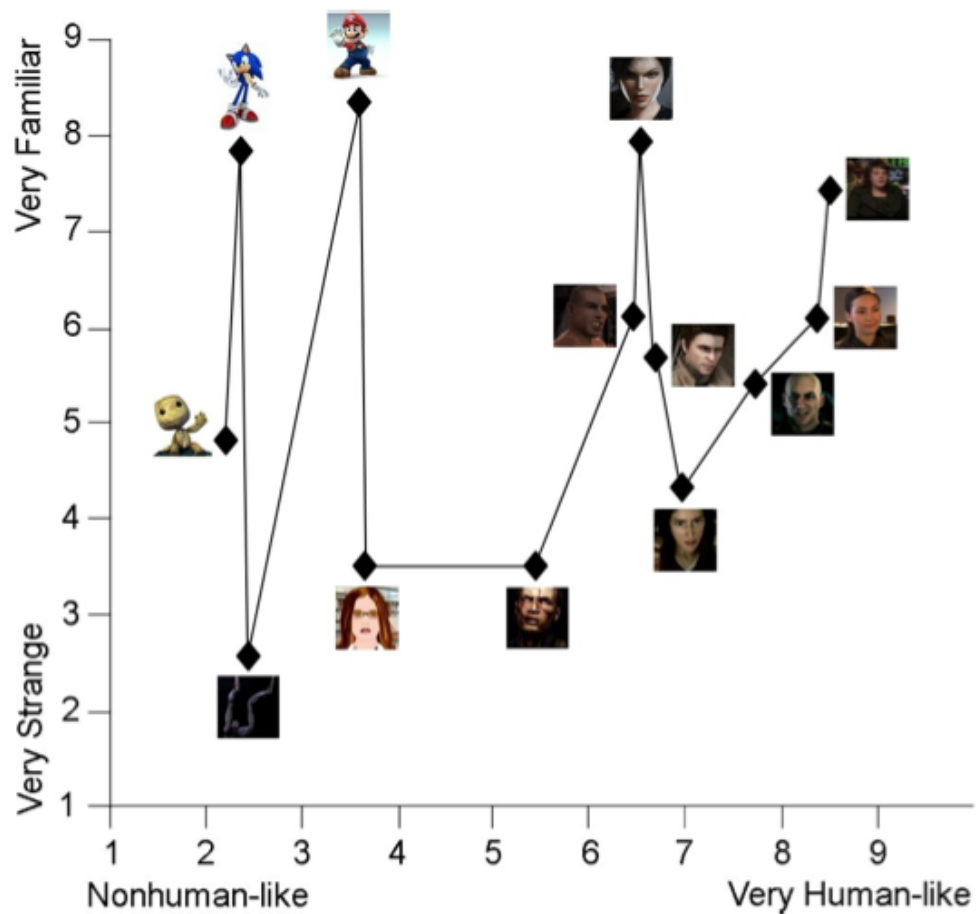


Figure 3: Tinwell (2009)'s graph showing three distinct valleys, which occurred when plotting mean human-likeness against familiarity.

A notable finding was that these valleys do not always represent the characters which were created to be the most disturbing. For example, the second valley contains two agents, the chat bot and the second zombie, with the chat bot rated as considerably less human and equally as strange as the zombie even though the latter has been designed to appear unsettling while the chatbot was designed for a communicative purpose and without the intention to disturb. The final, broader valley contains some characters designed to evoke fear but also the main protagonists from two popular games where there would be no such intention. An additional finding was that the genuine human character in the set was actually rated as less acceptable than the photorealistic characters which may have been due to the experiment being set in the context of video

games meaning that the participants were mainly experienced gamers who may have developed expectations for what would be acceptable in a video game context so the presence of a real human may have seemed incongruous. To explore the reasons why the acceptability of the characters did not follow the rise-and-dip pattern predicted in the classic UV graph, the researchers reviewed the characters that had been found to be the most eerie and noted that some idiosyncrasies of their animation may have contributed to these ratings. For example, there were subtle mismatches between the speech and the lip synchronisation which may have been unsettling. Therefore, the multiple valleys may arise from specific mismatches in the believability of the characters where a photorealistic appearance is paired with a speech performance that does not deliver the same level of verisimilitude.

The researchers went on to develop this theoretical framework further in a study (Tinwell et al, 2011b) where the lack of expression in different parts of the face was paired with different types of emotion being expressed. The intention was to look for a differential effect of a lack of expression depending on the emotion being portrayed. New video clips were created for this study to control the type of emotion being communicated by the agent. CG characters were created using a system where emotional expressions could be finely manipulated to control the magnitude of the emotion being expressed. The emotions that were tested were based on Ekman's (1992) core six of anger, disgust, fear, happiness, sadness and surprise plus a neutral expression. The controls allowed independent manipulation of the muscles in the upper and lower part of the face so it was possible to create videos where the expression was being displayed fully or was lacking in each region. A human actor was also filmed posing the same expression but given that these were natural rather than CG expressions, there was no corresponding lack of animation in the upper part of the face. Three conditions were

tested in the study; a human; an animation with fully articulated expressions; and an animation without expression in the top part of the face. A repeated measures design was used where videos from all three conditions were presented to the participants. This would have meant that participants had prior experience of the stimuli when viewing later clips in the series. Ratings of the human-likeness and strangeness of each character were again collected. An emotion identification task was also carried out where participants were asked which emotion would best describe the character. As expected, the lack of expression in the upper part of the face did cause an uncanny effect (and so the animated characters without expression were found to be more uncanny than either the fully animated characters or the human actor) but they found that the magnitude of this effect was modulated by the type of expression being communicated with a lack of expression of disgust, fear, sadness and surprise exaggerating a sense of the uncanny. However, there was no difference in uncanniness when the expression displaying a lack was either anger or happiness. It may be that the mouth shape conveying a smile is enough to denote happiness or a grimace for anger and so the information present in the upper part of the face is less relevant. Tinwell et al suggested that the cause of this unsettling effect was due to ineffective communication of the desired expression as a result of suppressed emotion in the upper part of the face, in other words, an inability to communicate expression fluently may function as a cue that all is not right with the person in front of us.

To explore one explanation for this reaction, another paper to be considered in this part of the review linked a lack of emotional fluency with a tendency to make inferences about the emotional or psychological state of the agent with which we are interacting. Many such inferences might be used as explanations for an abnormal expressive pattern but this study specifically tested the hypothesis that this is due to perceptions of

psychopathy. Tinwell et al (2013) observed that it is characteristic of people who score high on tests of psychopathy to not show a typical wide-eyed expression in response to startling or surprising stimuli. As her previous work (Tinwell et al, 2011b) had found a lack of emotion in the upper part of the face was important for the UVE, it was hypothesised that uncanniness was due to unconscious perceptions of this trait and the possible expectation of corresponding anti-social behaviour. They used the three sets of stimuli as described above: human videos; fully animated CG characters; and CG characters with a lack of animation in the upper part of the face. A scene for each character was filmed or constructed to represent the character responding to the noise of a scream, selected as a stimulus that participants should expect to startle the character. Male and female versions were created for each of the three types of character to explore the effect of character gender on the perceived uncanniness and psychopathic traits. Measurements of human-likeness and familiarity were collected as before but for this study, measurements of several negative traits were also collected including some which are specifically indicative of psychopathic traits along with more general negative evaluations. The results demonstrated that the characters lacking in animation in the top part of the face were again found to be the most unsettling and the extent to which participants perceived the psychopathic traits over the other negative traits was able to explain the variance in the uncanniness ratings. This relationship was particularly strong for the male characters and the authors suggested that this finding links back to the original hypothesis as males are more likely than females to display psychopathic and anti-social traits. The authors acknowledge that there are other cues which may be notable in causing a perception of the uncanny but the evidence in these studies suggests that an expectation of negative social outcome caused by the lack of expression in a certain part of the face can be seen as explanatory for the uncanny response to some virtual

characters. This research approach has been informative during the design of the present project and some elements of the method will be considered further in the Methodology chapter and introductions to Phase 3 of this project.

Finally, Creed and Beale (2008) explored mismatched emotional expressions. While this study does not explicitly measure human-likeness or reference the UV, combining their findings with the lack of expression evidence above has allowed the present research project to develop a second novel area for enquiry. Creed and Beale asked how participants would respond when interacting with a virtual agent whose voice track and the expression carried mismatched emotions. One possible outcome could be that the two would be blended into a new emotional composite, another that there would be no clear emotion communicated, or alternatively that there might be a dominance from one 'track' over the other with the sound or visuals taking precedence. Four emotions were used for the two modes: happy; warm; neutral; and concerned. 24 conditions were tested combining static presentations of the emotions with videos where all possible combinations were presented. They found that the mismatched presentation did cause participants to evaluate the agent more negatively, particularly where the warm or happy expression was paired with a concerned voice. There was also a high level of confusion expressed, but considering the issue of cognitive load, it was found that there were no significant differences in the time taken to rate the expressions. If the mismatch had been placing a higher cognitive load on the participants, a longer time to rate the agent would have been expected but this was not found. The nature of the agent and ethical considerations constrained the emotions that could be presented meaning an exploration of a wider range of emotions is needed to explore questions relating to the eeriness dimension of the UV experience.

1.4: Summary and Review of UVE Literature

This literature review has considered four theoretical explanations, and seven areas of empirical research into the UVE, and has reviewed how researchers have used a range of different theoretical frameworks and concepts to inform their approaches to investigating the effect. By considering the theoretical approaches and evidence-based research separately, it has been possible to critically evaluate what could be known from this body of literature, and it was concluded that four aspects presented the most promising areas of enquiry to develop in the present thesis. These were the theories of categorical perception and the evidence for trust and empathy, anomalous features and cue mismatch effects.

The other perspectives and explanations were not chosen for exploration in the present thesis as they were limited or flawed in various ways, which will now be detailed. It was felt that evolutionary aesthetics was unable to offer a reasonable explanation for the UVE, as the premise of an evolutionary drive to avoid eerie near-humans would predict a far stronger emotional reaction than the subtle sense of unease that characterises the effect. The neuropsychological explanations proposed that particular patterns of activity in different areas of the brain would occur when stimuli varying in human-likeness were presented, but the evidence gathered in support of these theories were flawed in not linking levels of activity to how participants actually felt when encountering the different stimuli. In addition, single examples of stimuli were often used which meant that conclusions of whether a valley was present in the data are difficult to draw. When considering the theory of fears of one's own mortality, it was felt that two issues were caused by viewing the UVE from the TMT perspective. Firstly, the evidence for TMT's general applicability was limited, and secondly, it represented a level of

abstraction away from the experienced emotion of uncanniness which hampered its ability to explain the effect. Finally, no compelling evidence was found for the error sensitivity explanation and it could reasonably be argued that an increase in sensitivity may actually *decrease* any sense of eeriness as additional information may serve to reduce any ambiguity between cues that could have contributed to an unsettling effect.

From the four promising areas of enquiry, the one that most influenced the design of the present research project was that of cue mismatch but the issue of anomalous features was also highly influential. The present research project took the mismatch issue as one of the key research questions as it has been demonstrated that an element of the uncanny experience occurs when there is a mismatch in the information presented in a face and evidence now suggests that a lack of expression may cause eeriness because of unconscious expectations of threatening behaviour. The role of mismatched facial expressions as perceptual cues, which may enhance a sense of the uncanny, is a novel area for enquiry and that is the knowledge gap that the final phase of this study will cover. In reviewing the results, perception of categorical boundaries and the influence of trust and empathy were also considered and used to inform the final conclusions about the research findings.

1.5: Face Perception Literature

It was identified above that one explanation for the UVE is the presence of features that are anomalous to the overall face. This literature is introduced here as it is particularly relevant to Phase 2 of the current project. The relationship between perceiving parts of faces and perceiving the whole face has been extensively researched in psychology, and this section of the review will present a selection of literature in this

area to help frame specific research questions on whether NHFs follow the same processing patterns as natural human faces.

1.4.1 Introducing Analytic and Holistic Processing

Humans possess a seemingly effortless ability to recognise familiar faces and to extract a wide range of information from both familiar and unfamiliar faces. In everyday life, a brief glimpse of someone is enough for us to be able to ascertain their gender, race and age, to perceive how they are feeling, where their attention is currently directed and whether they have any direct intentions towards us (Kanwisher and Moscovitch, 2000). In the case of familiar faces, we are able to identify who they are and access any facts known about the person even though any single example is just one from a vast number of faces encountered over a lifetime. While this may seem automatic and effort-free, this is actually an achievement of exceptionally sophisticated processing as faces are complex animated stimuli which share a broadly similar appearance and which are routinely encountered under different lighting conditions, at different angles and which markedly change in their physical characteristics over time. Despite these challenges, faces are recognised quickly and generally with a high degree of accuracy (Bruce and Young, 2000). The question of how faces are perceived has attracted attention from a wide range of areas and face perception research has developed into a broad and cross-disciplinary field with contributions from neuropsychology and computer science, as well as cognitive, social, developmental and evolutionary psychologists (Hole and Bourne, 2010).

Humans are so well attuned to the importance of face-like stimuli that we are prone to perceiving a face-like appearance even when no face is present. Little et al (2011) provided a striking illustration of this as shown in Figure 4 below where several inanimate objects appear to have a face-like appearance.

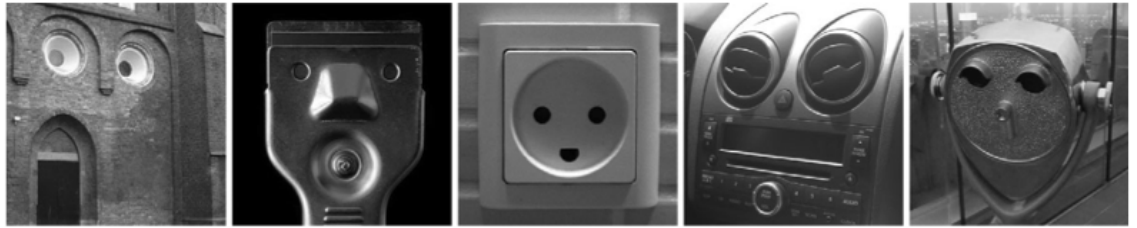


Figure 4: Non-face objects which are readily seen as face-like: from Little et al (2011).

[Images from Flickr users eworm, Aquario, mallol, Listener42 and vectr (provided under Creative Commons licences).]

The induced face-perception effect seen in the images in the figure results from the configuration of the elements of the objects as few of the components of those objects have any resemblance to the facial features they stand in for but nonetheless the face-like impression is compelling. To borrow a concept from Gestalt psychology, the whole impression is more than the sum of the individual parts. The question of this relationship between the part and the whole in how faces are processed has been at the heart of a debate for the last century with one consideration being the terms used to describe the ‘whole’ and ‘part’ elements. Peterson & Rhodes (2003) used the terms ‘analytic’ and ‘holistic’ to describe parts and whole respectively but noted that there was not a consensus on that definition, with the literature on face processing including such terms as ‘piecemeal’, ‘local’, ‘fine-grained’, ‘part-based’ or ‘componential’ as synonyms for their ‘analytic’ term and ‘global’, ‘configural’ or ‘coarse’ for ‘holistic’. This review will adopt the convention of using ‘analytic’ and ‘holistic’ unless otherwise noted.

One approach which has been taken to investigate the extent to which analytic and holistic processes contribute to how we perceive faces is to explore how that perception is affected when faces are presented upside-down rather than upright. The act of inverting a face changes the arrangement of features from that which we are used to seeing so that the standard configuration of mouth, nose and eyes is lost when the face is

turned upside down. As the studies to be presented below describe, faces are harder to recognise when they are inverted, leading to the suggestion that face perception is primarily driven by the perception of this relationship between the features rather than the features themselves.

This question was first posed by Yin (1969) who noted that faces represent a class of objects only commonly encountered in their upright orientation. This study explored the effect of how easily faces could be recognised if that orientation was changed, comparing this performance to the ability to recognise houses, another type of complex shape usually only seen in one orientation. The study tested four modes of presenting both faces and houses and participants learned to identify the objects and were then tested on their ability to recognise them. The faces or houses were presented in four different ways: upright when learned and when tested; inverted when learned and tested; learned upright and then tested inverted; and learned inverted then tested upright. The results demonstrated that both faces and houses were harder to recognise when inverted but faces showed the greater decrement in performance. This became known as the 'face-inversion effect' (FIE) and these initial findings established inversion as a marker of face-specific processes and also a tool for investigating what makes face perception special (Kanwisher and Moscovitch, 2000).

This review will describe selected publications about the FIE. Within face recognition research, evidence from three areas has been used to argue that faces represent a special class of visual stimuli which are processed in a unique way, and have tried to explore the reasons for this, which are the FIE, developmental studies and studies of the effects of prosopagnosia. Prosopagnosia will be considered in Section 1.5 under emotional responses to faces, so this section will cover developmental and behavioural evidence for the FIE. As it has been adopted in this study of the UVE as a tool for investigating whether

NHFs are processed using an analytic approach rather than holistically, the review of each study will be considered for its approach to investigating the FIE and, where relevant, evaluated in terms of the evidence supplied for analytic versus holistic processing.

1.4.2 Developmental approaches

Valentine (1988) posed the general question of whether the FIE suggests that faces are processed by a unique and specialised mechanism which is not shared in the processing of other complex stimuli. One of the strands for investigating this was to review developmental studies of the FIE to explore whether the effect was present from birth or whether it developed with age. An early sensitivity to orientation is observed in studies of infants between 16 and 22 weeks of age but there is also evidence that abstract patterns were similarly discriminated, with one study suggesting that this does not provide evidence of a face-specific sensitivity. Evidence for piecemeal encoding in very young children was found by Cohen and Cashon (2001) who explored whether seven-month old babies perceived faces as individual features or organised configurations. This study presented babies with images of two female faces followed by either images of familiar faces, images created by re-arranging familiar features into novel configurations (switched faces), or an entirely novel face. These were presented upright or inverted. They measured the amount of time babies spent looking at each type in each orientation. The results showed longer looking for switched faces in the upright condition but not when inverted, suggesting analytic processing for the inverted faces but holistic processing for the upright faces.

Returning to Valentine (1988), the developmental path of sensitivity to orientation is also subject to question: studies are reported that demonstrated that young children (aged under 10) did not experience a FIE. Instead it appeared at around 10 years of age

and it was suggested that young children process faces in a piecemeal fashion, rather than holistically, until the development of holistic face perception, which is marked by the appearance of the FIE. Diamond and Carey (1977) termed this change from analytic to holistic processing an 'encoding switch' and provided supporting evidence for piecemeal perception by demonstrating that children aged 6-10 relied on external cues such as sunglasses or hairstyles to identify faces but that this reliance stopped after age 10. However, Valentine (1988) argued that the design of this study was subject to floor effects preventing the actual inversion effect being measured and indeed further studies found evidence of a FIE in children of 7 years old. The idea of a change in encoding method during development remained open to question, and was examined in Brace et al (2001). This study used an innovative experimental design where the stimuli were presented in the context of illustrating a story rather than just as isolated images without narrative. This study also used an inverted versus upright recognition task where children (aged between 2 and 12) were asked to select a learned target face from eight distractor faces as quickly as possible. The inclusion of a large number of distractor faces was another innovation and was necessary to allow children across the age span to demonstrate their perceptual ability without risking floor or ceiling effects. The speed of recognition and recognition accuracy were measured. It was found that a FIE did emerge even in children aged 5-6, considerably younger than had been previously been observed, suggesting that an alternative to the encoding switch explanation for the observed pattern could be that the children were developing in perceptual expertise as they gained face experience. Unexpectedly, the youngest group of participants demonstrated an 'inverted inversion effect' where their recognition was faster for the inverted faces than for the upright faces. This effect could have been due to the presence of an encoding switch taking place at an earlier age than previously hypothesised or that the findings

could represent children's use of a generally inefficient encoding strategy, regardless of whether that was an analytic or holistic approach.

Finally, Pellicano and Rhodes (2003) compared the performance of child (aged 4 and 5) and adult participants on a test of recognition of facial features when presented in isolation compared to being presented as they would normally appear in a face. Participants were presented with a target face followed by two test faces. The test faces were either presented as a whole face or as just the eyes or mouth. The two whole test faces were the same except for the substitution of eyes or mouth from another individual; the two-part test faces contained face parts from different individuals. They found that, when presented upright, the components presented in the face were easier to recognise than the features in isolation and this applied generally across all the age groups. However, an inversion effect was found for all groups, with the decrement increasing with age. These results suggest that adults and children are able to use holistic processing for upright faces and that reliance on analytic processing may have emerged in previous studies as a result of the way that the tasks were constructed or the reactions measured.

1.4.3 FIE as a loss of configural representation

The present research project will include an experimental study using the FIE to explore how NHFs are processed. The classic explanation of an inversion decrement occurring due to a disruption of holistic processing will be considered first, followed by alternative explanations and interpretations. Farah and Tanaka (1995) tested whether the inversion effect occurred because of how the act of inverting something affects the whole shape rather than the detail of its individual parts and whether an inversion effect could also be induced in complex non-face stimuli. Their first study addressed the second

question using stimuli that were not in any way suggestive of faces but which had two face-like properties in that they were complex overall shapes composed up of smaller sub-shapes. These abstract stimuli were composed of patterns of dots which could either all be presented as black dots, or with groups of the dots presented in different colours. Where all the dots were black, this represented a holistic way to present the overall shape, with the differently coloured dots suggesting a part-based representation. The patterns were used as training stimuli for a recognition test where participants learned the shapes in the part or the whole representation and then were tested on their ability to identify them when presented as the whole shape. The results indicated that the inversion decrement was greatest when the shapes were learned in the whole representation rather than the part representation. Their second study used a similar approach but used actual faces rather than the dot-pattern representations. In this instance, the part-based stimuli were created by taking individual component features from sketches of faces (the eyes, nose, mouth and face outline with no features inside) and presenting them as the stimuli to be learned. The whole face sketch was used for the whole face conditions. The same pattern of responses was found as in the first experiment with the inversion decrement affecting the whole faces rather than those where the individual components had been learned. Their conclusion was that the FIE does occur as a result of faces comprising a complex whole which is rarely broken down into its component parts, but that other complex shapes could also be subject to a similar effect. However, the use of faces sketches rather than photographs as the stimuli in the second experiment may be a methodological concern in this study as it may be that the sketch representations are processed differently to photographs of faces.

In another exploration directly comparing the effect of manipulating individual features with the effect of manipulating of spatial relations, Searcy and Bartlett (1996)

created face stimuli (from photographs, avoiding the methodological concern noted above) which were manipulated to produce two types of new, distorted faces. This study is of particular interest to the current project as some of the distorted stimuli used in the study have definite uncanny qualities. In the first type of distortion, the individual features were altered with the goal of deliberately creating grotesque faces: eyes were given the appearance of cataracts or had the pupils tinted red, teeth were blackened, discoloured or lengthened to produce a fang-like appearance. In the second type, the features were moved within the face to create improbable and grotesque spatial arrangements. These altered faces were presented (upright and inverted) to participants who performed two tasks. Firstly, they rated the grotesqueness of each face on a scale of 1 to 7. They were then asked to perform a same/different classification when presented with pairs of the faces consisting of one unaltered face with either the same unaltered face or the componential or spatially altered face. The pairs were presented upright or inverted, and the experimenters measured both the speed of response to make a same-different judgement and also the percentage of participants who were able to make a decision in the allocated interval. The results of the grotesqueness ratings found that inversion reduced the grotesqueness for the spatially distorted faces but not for the componentially distorted faces. The results of the same-different task showed that inversion caused participants to find it harder to decide between the spatially distorted faces, with a fall of 44% in the response rate when these faces were inverted. These results both support the hypothesis that inversion interferes most with the holistic encoding of faces. Searcy and Bartlett's second study addressed a concern about the procedure for the first study that the 'different' pairs in the same-different task were always the original face plus a distorted face so the very presence of either type of distortion may have been available to participants as a cue to respond with a 'different'

response. To address this, they constructed new pairs where both faces were distorted. Two methods had been used to create each type of distortion so it was possible to create 'different' pairs of images of the same face where, for example, the eyes were blurred into cataracts and the teeth blackened in one image and the features moved to implausible locations in the second. This negated the presence of distortion alone as a helpful cue for participants who were asked to compare the two images. The original faces were never seen in this experiment. Again, the faces were rated for grotesqueness and a same-different judgement task used. The response patterns replicated those in the first study, suggesting that participants were not using grotesqueness as a cue and that the spatially distorted faces were subject to an inversion effect whereas the componentially distorted faces were not. These two studies provide evidence to support the role of holistic information in face processing.

Freire et al (2000) introduced an exploration of a delay in the time between learning and recalling a face when upright or inverted to test the hypothesis that the FIE occurs due to a 'bottleneck' in the encoding of holistic information. As with the study above, they compared componentially and spatially manipulated faces but without such a focus on creating grotesque faces. Instead the manipulations were more subtle and at first glance appear to be normal photographic portraits. Across two initial experiments they first found the expected pattern that inversion disrupted the recognition of the spatially altered faces but not the componentially altered faces. To explore the bottleneck hypothesis they introduced a memory task where participants were asked to view a target face and then pick it out from a pair of faces after either one second, five seconds or ten seconds. They found that there was an inversion effect on memory for faces that were spatially altered but none for the componentially altered faces, and that there was no significant effect of delay. This finding supported their suggestion that a bottleneck

occurs at the encoding stage of face representations and that this is responsible for the FIE.

1.4.4 Alternative explanations for the FIE

As has been seen in the studies detailed above, there exists strong evidence for the role of holistic representations in the FIE. However, alternative explanations have been proposed which challenge that explanation of the FIE and may suggest that other processes may have a role.

Firstly, Rhodes et al (1993) challenged the assumption that inversion increases recognition difficulty by disrupting holistic information. They found a large decrement in detecting changes to eyes and mouth, but not when the features were presented out of a face context, suggesting that this is due to encoding them in a face differently than presented in isolation outside a face. They suggest that this ambiguity in knowing when features and configuration have been disrupted imply that the picture is more complex than simply an analytic versus holistic notion.

Maurer et al (2002) suggested that a simple concept of holistic processing is insufficient and that rather than considering this as a single process, they identified three separate sub-processes. Firstly, the perception of first-order relations indicating a face is present (i.e., the detection of eyes, nose and a mouth) followed by holistic processing to combine the features into a gestalt, and then finally the processing of second-order relations to encode the spacing between the individual features. The authors present evidence from behavioural studies to suggest that inversion affects all of these sub-processes, and suggests that the prevalent view of holistic processing as a single process could benefit from consideration of these sub-processes.

Sekuler et al (2004) suggested that the inversion effect is not caused by a specific type of encoding but is instead a result of the amount of information available to an observer when viewing faces under different conditions. This study used a novel response classification technique to identify the regions of the face that were used to differentiate between images of different faces. While they did find a classic inversion effect in discrimination performance with the inverted faces being harder to recognise, they were able to analyse the areas of the face that were being used to make those discriminations and found that they were the same in upright as in inverted faces. Their conclusions were that inversion interferes with the efficiency of extracting information from faces and there is a quantitative difference but not a qualitative one between the processing of inverted faces.

Continuing the theme of examining the method by which information is extracted from faces, Hills et al (2011) examined the importance of fixation in understanding how upright and inverted faces are processed. Across three experiments they tested a 'feature-saliency' hypothesis by firstly carrying out a recognition test with upright and inverted faces, cueing participants to fixate either on the eyes or the mouth region by highlighting those areas with a cross. They found that drawing attention to the eyes reduced the inversion decrement on recognition accuracy compared to drawing attention to the mouth or not providing a cue at all. In a second study, they addressed a limitation of the first study where it was observed that the location of the cross may have served as a cue to where the top of the face was located. To address this the cross remained in a fixed location at the centre of the screen and the face was presented lower or higher to align the eyes or mouth with the cross, thus removing the salience of the cross as a marker of where to direct attention. These results were consistent with the first study, suggesting that the additional expectancy caused by the position of the fixation cross

could not explain the effects that were found. A final study tested whether giving the cue at the point of learning the face or being tested on the face would significantly affect the outcome. It was again found that the eye cue had more of an effect on the inversion decrement than the mouth cue and this occurred regardless of when the cue was used. When considering the UVE, if it is the case that anomalies in the eye regions of NHFs are acting as a cue to pay particular attention to this region, a decreased inversion effect may be found for those faces compared to human or artificial faces.

Finally, Schwaninger et al (2009) challenged the absolute nature of the holistic versus analytic explanations and instead suggested a dual mode hypothesis for face processing, where both types of encoding are fundamental to a fully processed perception of a face. A novel approach to separating holistic and analytic processing involved manipulating face photographs to preserve feature-based information but losing the configuration of the face (cutting faces into individual features and 'scrambling' them randomly) or to preserve the arrangement of the features while removing the detail of the features (blurring the face image to the point where the features were indistinct). They used these stimuli in three behavioural experiments and three computer modelling studies. The behavioural experiments explored whether it would be possible to recognise faces in conditions where only the features or only the configuration was available. This was found to be possible for unfamiliar faces as well as familiar faces. Those studies also measured the extent to which seeing one of the face types facilitated the perception of the other and the results indicated that these are separate and distinct processes. Their third study investigated the role of first and second order relational information about the features. As above, first-order relational information gives the position of features relative to one another and the second order relations are the specific distances between them. The scrambling technique used served to decompose the face into individual

pieces, randomly scattered, and so disrupted both types of relational information. In a final behavioural study, the face pieces were 'exploded' rather than scrambled meaning that they were presented in the original arrangement but with the spacing altered to disrupt the second order relational information. In this way it was possible to measure the extent to which this disruption affected a participant's ability to recognise the faces and the results demonstrated they found that the recognition sensitivity for these faces was not significantly different from the fully scrambled faces, suggesting that the second-order relational information between the features is not used in the encoding of faces into memories.

This section has reviewed the FIE and considered evidence suggesting that it arises as a result of disrupted holistic processing as well as explanations suggesting that there may be other factors at work in the observed degraded performance when faces are inverted. Several common techniques used to measure the FIE have also been detailed in this section. A question arising from the earlier review of the literature specific to the UV was that of how NHFs are processed was whether NHFs are processed in the same way as human faces. Through studies of the effect of anomalous features it had been observed that uncanny sensations could emerge when eyes were distorted or where there was a mismatch between different components of the face. This leads on to a question of whether analytic processing is being engaged for NHFs to a greater extent than is usually engaged for human faces. The FIE effect will be adopted as a tool for investigating this question and drawing conclusions about differences in how human, near-human and non-human faces are processed.

1.6: Emotional Expression Literature

This review has considered the specific issue of the FIE which will be investigated in Phase 2 of the present project: Phase 3 continued to explore face perception but focused on facial expressions and the potential role of mismatched expressions in eliciting a sense of the uncanny. This section of Chapter One begins with a general background to the psychology of facial expressions, followed by a review of the literature on emotional expressions in faces which have been CG or otherwise manipulated leading on to a discussion of mismatched or incongruous expressions. Finally, to build on the considerations identified in the second phase of this project, the effect of inversion on perceiving emotion from faces will be considered.

1.5.1 Processing expressions of emotion

As described above, humans have sophisticated expertise in extracting information from faces under widely varying circumstances. One key source of information to be extracted, processed and used is an understanding of how other people are feeling as being able to see how an interlocutor is feeling allows us to predict their behaviour and tailor our own response accordingly. In terms of the UVE literature, Canamero (2006) initially suggested an incongruence between the upper and lower halves of the face may have a role in eliciting an UVE as it could affect the ability to perceive emotion from the face. A fundamental question in this area asks what is the nature of the core emotions that can be communicated through facial expressions. Ekman and Friesen's (1978) classification of emotional expressions is one route for addressing this question. Their approach is known as the 'facial action coding system' and it provided an extremely detailed method for both codifying the emotions that could be conveyed through facial expressions and describing the muscular mechanics that elicit each expression. While

over two hundred different expressions were detailed in the full classification, Ekman (1992) also made a case for six emotions to be recognised as 'basic' by virtue of their universal nature, distinctive causes and emotional responses and also the timeframe over which they occurred. These six were: anger; fear; sadness; enjoyment; disgust; and surprise. Contempt, shame, guilt and embarrassment were also considered but were not considered to meet the full criteria for basic emotion status. While the universality of these emotions became the subject of intense debate (Russell, 1994), their distinctiveness and links to emotional response mean they have become generally accepted as basic emotions and have been cited in previous research into the UV. As a result, these basic emotions will be used later in the present project both in creating stimulus materials and also as response items for participants to use in identifying and labelling emotions.

However, research has found differences in the recognition performance for different emotions. While these may be considered basic emotions, they are not all recognised with the same speed and accuracy. Calvo and Lundqvist (2008) compared the recognition performance for faces posing neutrality, happiness, anger, sadness, surprise, fear, and disgust and found that while neutrality and happiness were identified very quickly and accurately, there were areas of confusion with significant differences in the recognition performance for negative emotions. Two symmetrically crossed pairs of emotions emerged, these being anger-disgust and surprise-fear, with sadness misclassified as both disgust and fear. The happiness superiority effect had been found in previous studies, but this work particularly emphasised the differences between different negative emotions. This is particularly salient for the current consideration of the UV as the experience is characterised by a negative emotional experience.

1.5.2 Experimental Studies of Emotional Expression

Having considered general issues relating to emotional expressions, this review will now consider specific studies concentrating on recognising emotion in faces where some form of manipulation has been used to create the stimuli.

An early study in this area used images of emotional expressions created by morphing and investigated whether these would be perceived in the same way as 'natural' photographs of emotional faces. Takehara and Suzuki (1997) used morphing software blending pairs of images of emotional expressions, blending happiness into surprise, surprise to fear, fear to anger, anger to disgust, disgust to sadness, and sadness to happiness. Participants were shown the original images and the intermediate combination faces and asked to rate them on happiness, sadness, anger, fear, disgust, and surprise. Prior to the experiment it was not known whether the categorisation of the intermediate expressions would follow the same patterns as the processing of natural expressions. Takehara and Suzuki visualised the results using a two-dimensional scaling model of pleasantness versus arousal and plotted the location of each face combination. An examination of this plot shows similar patterns of confusion between fear-surprise and anger-disgust to those observed by Calvo and Lundqvist (2008), suggesting that these combinations do share common elements. However, the main conclusion drawn from the research was that the morphed faces were perceived in a similar fashion to the natural emotional expression faces as they were rated in the expected order despite the ambiguity produced in the intermediate expressions. These findings suggest that morphed images are understood as valid psychological stimuli, and that emotional expressions can still be perceived under ambiguous circumstances.

The concept of categorical perception has been considered with reference to the UV earlier in this review. Calder (1996) explored the categorical perception of morphed facial

expressions, as it had been previously found that line drawings of emotional expressions were perceived categorically but there was uncertainty as to whether this would also apply to morphed expressions. As with the study above, morphs were created by merging images of one intentionally posed expression into a different expression: in this study, the expression pairs were happiness to sadness, sadness to anger and anger to fear, examined in two studies. They found evidence for categorical perception for all three of the pairs. However, they observed that this effect may have occurred as an artefact of the way the studies were constructed with a disproportionate presentation of the morph images compared to the anchoring original images. To address this, a separate set of images were selected from a single model representing a continuum ranging from happiness through anger to fear. These images were used in two studies of recognition and identification and categorical perception effects were still found. Developing this idea, Calder et al (1997) explored caricaturing which is another type of face manipulation technique where the positions of features can be altered with a resulting change in appearance. In the case of emotions, the physical distances between features in a neutral expression and a specific emotional expression can be increased to produce a new and often more emphatic presentation of the target emotion. Conversely, decreasing the distances creates an anti-caricature. The researchers hypothesised that the caricatured images would be recognised faster than the original expressions, and the anti-caricatures would be recognised more slowly than the originals. Their findings supported this hypothesis for all six basic emotions and caricatures of each emotion were recognised faster than the original faces which were in turn recognised faster than their anti-caricatures. Following on from this finding, a second study explored whether this would also hold when the caricatures were created using a different starting point. The inexpressive neutral face noted above was replaced with an 'average' face created by

blending together all six basic emotions. This was to understand more about the underlying mechanism involved in the caricature effect. If the effect occurred as a result of perceiving changes of muscles when an expression changed, the neutral face-based caricatures would differ from the average or prototypical-based caricatures. The results of the second study found that both types of caricature produced the pattern of results found in the first study that caricatures improves performance while anti-caricatures decreased performance. An explanation proposed for this finding was that the manipulations represented the types of changes that might be found as emotions increase or decrease in intensity, thus how strongly recognisable they appear.

The part of the face responsible for conveying different emotions has also been examined. It has been found that some emotional expressions are recognised mainly from the top part of the face, some from the bottom part of the face and some seemingly unaffected by top versus bottom. Hanawalt (1944) found that the pleasant emotions (happiness, warmth and mirth) could be easily perceived from the mouth or lower part of the face, but less pleasant emotions such as fear and surprise were perceived mainly from the eyes or the upper part of the face. Bassili (1979) found that there was a difference in the emotions accurately recognised in the top and bottom half of the face, with happiness and disgust recognised from the bottom of the face, anger and fear recognised from the top half and little difference found for surprise and sadness. In addition, Baron-Cohen et al (1997) found that even complex emotional states such as scheming or guilt could be recognised from eye-expressions presented in isolation from the rest of the face. More recently, this has been extensively researched using composite faces, which represent another method for creating faces that cannot be encountered in the course of everyday life. The technique joins parts of faces together to form a new composite face, commonly formed by dividing a face across the bridge of the nose and pairing a top part

of forehead and eyes with a bottom half of nostrils, mouth and chin. Composites of different people can be used to investigate which parts of the face are used in recognising identity, and composites of the same person intentionally displaying different emotions can be used to investigate how emotions are recognised.

Calder et al (2000) used composite faces in four studies investigating emotion identification. The composites combined two pictures of the same model posing different expressions by using a top-recognisable emotion in the top part of the face with a bottom-recognisable emotion in the bottom part. It had previously been found by Young et al (1987) that participants were slower to identify a face created from a composite of two different people, and it was expected that recognition of emotion would also be slower from a composite face where the top half was displaying a different emotion to the bottom half. Calder et al's first study did find this effect as participants who saw the composite faces were slower to identify the emotion compared to those who saw the two halves mis-aligned. More mistakes were also made when recognising emotions displayed as part of a composite. The fact that the slower recognition and greater error rate (together known as the composite effect) did not occur when the faces were mis-aligned suggested that the ambiguity caused by the composite face was a product of the holistic combination of the two halves rather than an intrinsic property of seeing two emotions together. In a second study Calder et al also found that the composite effect did not occur when the faces were presented inverted. The role of inversion in disrupting perception of expression will be considered in more detail later in this review. To verify that the composite effect was not due to a novelty effect of the compositing technique they tested composites created from different models and found that no effect occurred. It can be concluded that creating composite faces by combining different emotional expressions makes it harder for those emotions to be recognised. It is

theorised that this may link to the UV and be a potential explanation for why some NHFs may appear unsettling. These findings will be considered in detail during Phase 3 of the present project.

The links between specific arrangements of facial features and perception of emotional expressions was investigated by Tipples (2007) in two experiments looking at how a threatening facial expression could be enhanced by the presence of an open mouth and widened eyes. The study was designed to extend existing research by Lundqvist et al (1999, 2004) which had suggested that a threatening expression could be signalled just by a V-shaped eyebrow. It had been suggested that this was because that arrangement of features formed an angular shape which was associated with unpleasant consequences. Tipples (2007) aimed to build on previous findings by adding the two new attributes and also by using more detailed face renderings than had been previously used. In their first study, CG images were created from a standard male base face to cover every possible combination of eyebrow shape, mouth shape, mouth curvature and eye wideness. Participants rated each image on how pleasant it appeared, how aroused the person appeared to be and how threatening the person seemed. Their second experiment added a female model as well as the original male model. In both studies, the ratings of pleasantness were lower and the ratings of arousal and threat were higher when the eyes were wide and the mouth was open compared to normal eyes and a closed mouth.

1.5.3 Concealment and deception: the effects of partially-seen expressions

The studies reviewed so far have considered the perception of emotional expressions under circumstances where those displaying the emotion are depicting a genuine feeling, either as a ‘candid’ naturalistic emotion or one deliberately posed to create a ‘happy’,

'angry' or similar expression. This review of emotional expressions sits within the context of research into the UV, and the starting premise for its inclusion is how emotion perception may be affected under non-normal circumstances. This review will now turn to research conducted on the effects of expressions which are concealed, partially seen or masked either deliberately or in response to social cues about the emotions that are socially acceptable to display.

The first to be considered here is that of genuine versus non-genuine smiles and the impressions formed when each are used. Within the mechanics of facial expressions, one of the areas considered by Ekman et al (2005) was that of the smile. A smiling expression may have a core role of indicating happiness, but Ekman explored the role of 'masking smiles' which are ones made when someone is lying. A masking smile is one in which the person is trying to conceal a negative emotion so that a smile might comprise a smiling mouth combined with 'leaked' facial movements from other negative expressions. In their study exploring this effect, participants were interviewed while watching either a disturbing or a neutral video and were asked to try to conceal their emotions. They were videotaped during this process, it is an ethical point of note that their consent to the video taping did not take place until after the recording had been made. Analyses of the video tapes showed that even though the mouth produced a smile, negative expressions were present in the rest of the facial expression. Ekman suggested that these 'leaked' negative expressions could give an impression of untrustworthiness where it is clear that the smiles are not genuine and that someone is trying to conceal something.

1.5.4 Expression and Inversion

The present project will consider the impact of inversion on perceiving emotion from 'blended' faces, which are similar to composite or chimeric faces but rather than taking

halves of faces as a whole, they blend a particular region of interest into a different base face. In the case of Phase 3 of this research project, these were created to blend emotional expressions so a happy base face could be shown with a sad, disgusted or fearful expressions from the eye region. The main aim of Phase 3 was to examine whether any of these blends produced eerie or unsettling responses in participants but following the findings from the Phase 2 of this project, the effect of inverting these blended faces was also studied to see which, if any, emotions would be recognised from the inverted versions of the face blends. The studies below have been included as a review of relevant literature on the effect of inversion on the processing of emotional expressions.

Firstly, Fallshore and Bartholow (2003) used schematic drawings of faces displaying surprise, sadness, anger, happiness, disgust and fear to explore whether these expressions would be recognised when the images were inverted. They found that sadness, anger, happiness and fear showed an inversion effect as participants were significantly less accurate at recognising these expressions when the faces were presented inverted. However, surprise and disgust showed no such inversion decrement as there was no significant difference in the identification of the emotions when the faces were inverted.

Prkachin (2003) reported studies of emotion labelling (the ability to give a linguistic label to an expression of emotion) and emotion detection (the ability to identify that a particular emotion is present) for inverted and upright faces. The emotion labelling study found that inversion impaired participant's ability to accurately label emotions and detailed how each emotion was affected by inversion. It was found that there was a relationship between how easily an emotion can be recognised when upright and the extent to which this is affected by inversion, with the faces harder to recognise when

upright being most affected by inversion. However, Prkachin questioned whether part of that pattern may have been due to the cognitive load imposed in the type of task used for the study as choosing which label to apply to a face when inverted can be a cognitively challenging task. To address this, for their second study participants simply had to state when they saw a particular target emotion on a displayed face rather than identify and label it. Inversion was found to disrupt sensitivity on this task. The pattern of particular emotions and how they were differentially affected by inversion on each task was considered. It was found that on both tasks, performance for labelling or identifying the expressions of anger, fear, happiness and surprise was similar but labelling disgust in particular was disproportionately affected by inversion. The response pattern to sadness was also interesting. While it was not difficult to label sadness when presented upright, identification of the sad expression was difficult and this was made harder when the face was inverted. It was found that participants would confuse sadness, anger and disgust when inverted. It was concluded that perceptual processing of emotions was affected by inversion, but the nature of that effect varied according to the emotion concerned.

However there has been debate as to the extent to which inversion affects the perception of emotional expressions. Lipp et al (2009) questioned the extent of the detrimental effect of inversion and found some responses to emotional content were augmented, rather than impaired, when faces were inverted. This meant that happy faces were processed as more positive and angry faces were found to be faster to detect.

The development of an inversion effect for emotional expression perception has also been considered. Durand et al (2007) explored the role of holistic information in the processing of emotion in two tasks involving adults and children of five different age groups. Their overall finding was that holistic information has the same role in emotion processing as for identity recognition, and that this information was being used by

children as young as six. They also identified the developmental path for emotions, as the ability to perceive different emotions emerged at different ages. Basic happiness and sadness were the first to emerge at 5 to 6 years of age and may represent a basic valence discrimination ability that appears as a precursor to the perception of more complex emotions. After this, the ability to detect fear emerged at around 7 to 8 years, anger around 9-10 and disgust around 11. A neutral expression was often read as a different emotion up to the age of 9. They concluded that this did not indicate that children cannot understand or experience those emotions, but more that they were unable to adequately predict the effect that an emotion would have on the facial expression of another person in such a way as to be able to identify it when seen. The developmental path and evidence of both inversion and composite effects found in this study mirrored that found in identity recognition and this study was the first to explore whether this effect was transferrable from identity to expression processing.

Finally, a different approach to the role of configural and holistic information in recognising emotions was taken by Martin et al (2012). The researchers observed that configurations of specific facial features could be used to identify particular emotional expressions, suggesting that there is a role for local processing of features in the recognition of those emotions. If this was the case, participants who were cued to use a local processing technique rather than a global one would have an advantage in recognising the emotional expression displayed. This paper reported the results of a study designed to explore this premise and found that local cueing did give participants accuracy and speed advantages in identifying all emotions when compared to a global orientation. This supports the importance of individual feature perception in decision-making about emotional expressions.

This final section of the literature review has presented a selected background to the perception of emotional expressions to give context to the final research phase of this project, the study of mismatched emotional expressions and their potential for explaining part of the UVE.

1.7: Research Design

This review has considered theoretical perspectives on the UVE and evidence collected from a range of different approaches. The current project was initiated during the period covered in Section 1.2 where ideas were beginning to be proposed about the UVE but there were no empirical studies which had demonstrated evidence for those ideas. At that stage, no empirical work had been produced which set out to operationalise what was meant by the ‘familiarity’ axis of Mori’s graph, to address ways of systematically varying the dimension of human-likeness or to map the emotional components of the UVE response itself. This research set out to do this with an initial exploration of the area. As those results were analysed, further research became available that allowed the key areas of anomalous facial features and cue mismatch to be identified as particularly salient. This allowed a focus to be developed on two specific novel research questions: whether NHFs are perceived differently from human and artificial faces, and whether mismatched facial expressions could provide an explanation for why the UVE appears for NHFs. The theoretical background to those research questions has been presented in this Introduction and Literature Review chapter. Chapter 2 will now describe the Methodological approaches which were chosen to answer them.

Chapter 2: Methodology

2.1: Theoretical Background

2.1.1 Structure of the Methodology Chapter

This thesis presents studies which have used psychological research techniques to understand the UVE in NHFs as a perceptual and a social phenomenon. This Chapter contains four sections which describe and justify the methods that have been used. Section 1 will describe the theoretical context in which the research has been grounded. Section 2 will detail the chosen methods and techniques for gathering participant data. The third section will present several methodological issues that are specific to the UVE and discuss how these have been addressed in the research studies. The final section will cover the ethical issues relating to research in this area and describe how these have been addressed while working with participants.

2.1.2 Theoretical Background

The research presented in this thesis uses the social cognitive approach to explore the features of NHFs which may cause unsettling or disturbing reactions in a viewer. An early and broad definition of social cognition was the study of how people think about people (Wegner and Vallacher, 1977). Carlston (2013) saw this as being formed through social cognitive events where impressions about other people are created or changed as a result of an interaction, leaving each party with new memories and impressions of the other. To adapt Kunda's (1999) definition, social cognition can be understood as the way in which cognition, motivation and affect (or thoughts, goals and feelings) are used to make sense of the social world to understand the people that are encountered in daily life. Interactions are shaped by pre-existing assumptions, prior knowledge and the social

context of the encounter and an understanding of the social world is shaped by what participants in encounters bring to those interactions in terms of beliefs, attitudes and ideas. The approach combines the methodologies of cognitive psychology with theories from social psychology to move beyond the quantification and measurement of social behaviours to explore the meanings those encounters have for the individuals involved.

This thesis is specifically concerned with social cognition as applied to face perception. It may seem self-evident that viewing a face-like stimulus will activate the perceptual mechanisms of decoding its identity and evaluating its familiarity along with judgements of the face owner's attractiveness, emotional state and intentions but face perception is a relatively new area for research within social cognition. Quadflieg and Macrae (2010) noted that although cognitive science and neuropsychology had made significant progress in mapping the processes and mechanisms of face perception, it was only recently that experimental social psychologists had started applying those theories and findings to the perception of people in a social context. However, this area of research is now gaining considerable attention and the importance of this interrelation between face perception and social evaluation is described by Hugenberg and Wilson (2013) as:

Faces are central to social cognition. They are informationally rich and processed with remarkable efficiency. Across multiple lines of converging evidence, it seems quite clear that face perception and our social cognitions about others are intimately linked. ... We see such recent advances at the interface of social cognition and face perception to be an exciting development, with social cognitive psychologists increasingly treating the face as a serious topic of study. (p187)

Similarly, Quinn and Macrae (2011) proposed an integrated social cognitive theory of face perception which could account for the social attributes that are perceived when seeing faces as well as extracting information on identity and emotional state.

This research aims to contribute to this growing area through an examination of human-likeness. Social cognition has traditionally been interested in interactions between

people, but encounters with NHAs are becoming more commonplace. Near-humans are encountered as virtual assistants or companion androids, or in multiplayer computer games where the other players are depicted by their human-like avatars. It therefore becomes relevant to consider the nature of these encounters and to consider those features of the NHAs that may make those encounters less than pleasant. The objective of the present research is to understand the relationship between the 'unusual' appearance of the NHA and the viewer's emotional response to it. This type of response is only relevant if the agent is considered in a social context with some potential for impacting on the viewer. To adapt Wegner and Vallacher's definition mentioned above, this research could be seen as exploring 'how people think (and feel) when someone is not quite human.'

2.2: General Methodological Issues

2.2.1 Research Approach

Phillips and Pugh (1999) identified three basic types of academic research: testing-out; problem-solving; and exploratory. In testing-out research, studies demonstrating an existing theory or finding would be identified and then replicated under different circumstances with the aim of testing the extent to which that theory could be considered generalisable and mapping the limitations of the existing theory. Problem-solving research applied academic methods and scrutiny to real-world problems which often require the application of multiple techniques and methods. It has the goal of arriving at a solution to the problem as well as an academic piece of research. Phillips and Pugh noted that this could be problematic as the drive to find and implement a practical solution could overwhelm the research aspects of the project. Both testing-out and problem-solving research were based on an existing theory or problem. In contrast,

exploratory research involved tackling questions about which very little was already known and where the researcher would need to begin from first principles in establishing their research concepts, developing new theories where appropriate and deciding which methods to apply to investigate the subject. In terms of Phillips and Pugh's classification, this thesis can be seen as an exploratory research project as there was no existing framework for social cognitive psychology research into the UVE and no methodological tradition to build upon. Instead, early findings about the UVE in disciplines such as human-computer interaction or robotics were used to design a suite of studies, beginning with collecting thoughts and feeling about NHAs through to two controlled studies which tested hypotheses developed from the results of the qualitative study.

2.2.2: Survey Research

'Wherever we need to know the characteristics of a population, or its resources, or its needs, or its opinions, the natural thing to do is to go out and ask questions.'
(Sapsford (2006), p4)

This study used surveys in Phase 1 to collect thoughts and feelings about NHAs. This was one of several methods that could have been chosen this data collection, where it may have been equally valid to carry out interviews or observations of participants encountering NHAs. Sapsford (2006) describes five questions for assessing whether a survey is the appropriate method for any research question, the first of which is whether it is feasible to research the topic at all, to eliminate topics with unanswerable questions or a population that cannot be identified. The UVE certainly can be researched as its scope touches on many areas of human enquiry and the effect is assumed to be broadly applicable to the general population so it is possible to draw conclusions from the results given by a sample of participants.

Secondly, can survey research answer the questions that need to be asked? Survey questions are fundamentally a means of prompting participants appropriately to elicit useful responses. A survey would not be appropriate if the topic could not be formed into discrete questions or if they were too ambiguous to answer clearly. The UVE had been clearly described as a phenomenon in the literature in such a way as to make it possible to formulate questions about the NHAs thought to trigger the effect, and to provide examples of those agents as stimuli for participants to respond to.

Thirdly, if it is possible to ask questions about a topic, can those answers provide valid conclusions? This condition requires the use of appropriate question types to collect measurements at a level and granularity that, when analysed descriptively or inferentially, are meaningful in terms of the research question. Satisfying this for the UVE meant adopting some measurement scales that had been used in previous studies (e.g. the eeriness, strangeness and human-likeness scales first used by MacDorman (2006)) as well as designing new questions and measures (e.g. the categorisation tasks and emotional expression ratings used in the Phase 3 and the scenario-based description and emotion tasks used in Phase 1).

Fourth, Sapsford (2006) asked is it ethically appropriate to survey participants rather than use any other method? One advantage of surveys, particularly those completed online, is that they can be completed privately and anonymously but for some topic areas this distance may actually be a disadvantage. For example, if the subject matter is potentially upsetting then a method where the researcher has more of a presence and is able to interact with the participant might be ethically preferable. This aspect was carefully considered as the *potential* to unsettle or unnerve is at the heart of the UVE so at least some of the experimental stimuli needed to be capable of potentially causing disquiet in participants. This issue is discussed more fully in Section 2.4: Ethical

Considerations below but it was concluded that the risk of disquiet did not present enough of a threat to participant well-being to necessitate a face-to-face presence for all phases of research. Instead, several methods for making contact with the researcher were always available for participants taking the surveys online.

Finally, is any type of research ethically and politically appropriate, given the nature of the research questions? This broad and philosophical question is perhaps most applicable to applied research where the conclusions may influence economic or social policies or treatment decisions. These applications are unlikely for research into the UVE, where the conclusions are likely to be used to inform future research into this specific topic or broader work on face perception and potentially to be applied to animation or computer graphics developments which are unlikely to present an ethical or political barrier to research in this area.

It was concluded that surveys were an appropriate and valid method for collecting data on the UVE, supplemented with results from a more controlled study to be described below.

2.2.3: Collecting Data Online

With the exception of a laptop-based experiment in Phase 2, all data collection for the present research was collected online, mainly by directing volunteer participants to a web page where the studies were hosted. Online research methods are a fast and effective way to systematically present stimuli to large numbers of participants (potentially allowing many simultaneous completions at once) and collect their responses without requiring any intervention from the researcher once the study has been designed. It is an efficient method for data collection at scale and can be considerably more cost-effective compared to postal or laboratory-based data collection. (After Hewson (2003).)

Reips' (2007) guide to carrying out online experiments identified eighteen separate benefits of conducting research in this way. As well as the efficiency and cost-saving benefits noted above, another relevant benefit is that online research is truly voluntary as participants cannot be compelled to visit a particular website and take part in a piece of research. Online research also benefits from a reduction in demand characteristics and experimenter effects as the researcher is able to essentially 'step out of the picture' after the web study has been introduced and allow the participant to work directly with the stimulus materials. One issue which is often raised with reference to online research is the extent to which findings can be generalised to the wider population, especially if a simulation of real-world behaviour is used during the study. Reips cites Horswill and Coster's (2001) study of driver risk-taking as evidence of the generalisability of online research. Horswill and Coster found that the mode of data collection was not a significant confounding factor and that the results collected online were as sensitive a measure of risk-taking behaviour as those collected using more traditional methods.

Reips' guide gives a comprehensive overview of issues to be aware of in online research. Five of these were identified for attention in designing the present studies and processing the results. Firstly, drop-out or non-completion is a significant concern with online studies as participants can stop at any point at little cost to themselves. They may lose interest, be interrupted or experience technical difficulties and unless there is a compelling reason for them to return and carry on, this results in an incomplete response. Reips recommends techniques for addressing this such as ensuring that participants have a realistic estimate of how long the exercise will take before they start, ensuring that page loading times are kept to a minimum to avoid boredom and frustration and ensuring that participants are able to return to complete the study if they do step out. These were implemented in each of the present studies. It was also desirable to avoid any invalid

access to the study as if online studies are included in search engines it can be possible for them to be 'completed' by tools that are attempting to index the site content which do not make the distinction that it is a data collection page rather than a standard web page. This was avoided by using software tools that blocked search engines from accessing the web pages which also prevented multiple submissions by the same participant. Reips discussed issues of recruitment in detail, exploring the benefits of mailing lists, participant panels and word-of-mouth recruitment and cautioning against distributing online studies to acquaintances or family. These principles were followed in the design of all studies. Finally, care should still be taken to ensure that no particular demographic groups are overly represented or excluded from the research, particularly if there is a reason to think that certain types of participant may respond systematically differently. Hewson (2003) recommends collecting demographic details about participants to be able to monitor any relevant characteristics. In the present research, gender, age and location were identified as participant-specific variables that may influence susceptibility to the UVE either directly or as a proxy for experience with technology and therefore levels of comfort with NHAs.

On balance, the online method was found to be an effective way to gather the data that was required for this exploration of the UVE.

2.2.4: Coding Qualitative Data

Phase 1 included two 'imagined encounter' tasks where participants were asked to view an image of a face and give their answers to two open questions asking how they would describe it to someone else and how they would feel if that face was used for a domestic robot living in their home. An open question approach was taken in this exploratory phase to avoid biased or leading language and to capture descriptions or

emotions that might not necessarily fit within a taken-for-granted taxonomy of uncanniness. This approach was chosen to elicit detailed and rich responses from participants in their own words. These narrative comments needed to be coded to learn more about the approaches that participants took to answering the questions, the physical features they chose to emphasise in the descriptions and the emotional responses they reported at the idea of sharing their home with the NHAs. Basit's (2003) review of manual versus electronic coding provided a useful overview of techniques for how coding frames can be implemented and an electronic method was chosen as a result. The process of transforming these narrative comments into coded values was informed by Swift's (2006) three types of coding system; representational coding, anchored coding and hypothesis-guided coding. Each system assigns text items to one or more categories but the methods are subtly different. Representational coding uses the content of the text to decide whether it should belong to a particular category and is often based on keywords so if a participant uses a key trigger word, the comment will be categorised as relating to that theme. Anchored coding develops this by only including categories that are directly relevant to that question that had been asked so that some of the codes that may emerge from a representational coding could relate to other questions in the survey so text referencing those would be classified as 'other'. Finally, hypothesis-driven coding uses the principles behind the design of the instrument to create a list of relevant codes to test each piece of text against and categorise accordingly. This latter hypothesis-guided system was used for the imagined encounter responses as previous research on holistic and analytic face processing informed the categorisation of features in those terms and the classic UV hypothesis that NHFs may be judged as eerie or unsettling was used to design a coding for positive, neutral and negative emotion terms. Sapsford (2006) observed the unscientific nature of this process as the coding frame itself is a subjective

product of the researcher's concepts relating to the subject matter, and although efforts were made to ensure that the frame was clear and unambiguous, it is inevitable that there is a subjective element to deciding how to code each phrase.

2.2.5: Controlled Experimental Study

Phase 2 and 3 were designed as controlled experiments, testing hypotheses generated from the findings of Phase 1 and grounded in the UVE and face perception literatures. A controlled experiment is defined in Clark-Carter's (1997) handbook of quantitative research as a situation where:

'The experimenter manipulates an aspect of the situation and measures what are presumed to be the consequences of those manipulations... It is felt that the properly designed experiment is the best way to identify causal relationships. By properly designed, I mean one in which all those aspects of the situation which may be relevant are being controlled for in some way.' (p5)

Study 4 used the FIE to investigate whether participants responded differently to NHFs compared to human and non-human faces and to explore the nature of that difference. Study 5 explored whether mismatched faces would elicit eeriness, evoke strong emotion or be difficult to identify or categorise. The experimental method was chosen to ensure the external and internal validity of the studies and demonstrate that any differences in response could be confidently attributed to the differences between the stimulus images. Clark-Carter details three threats to external validity and twelve threats to internal validity which should be taken into account in order to have that level of confidence in the findings. To ensure a study is externally valid and can be meaningfully generalised beyond the sample population, it is necessary to design the study in such a way as to control for those particular aspects of the participants and the task that could introduce bias into the results. It is assumed that the UVE is a phenomenon that should be broadly found in the general population so it was important

to ensure that the participants did not over-represent any particularly demographic group and this was achieved both Study 4 and 5 by randomly allocating participants to experimental groups. In terms of the tasks themselves itself, Clark-Carter suggests ensuring that design choices on aspects such as length of presentation and image type are considered for their ecological validity so that as far as possible, they should replicate experiences that participants may come across outside an experimental setting. This was accommodated by ensuring that the images were high-definition where possible and presented in colour.

Several of the cited threats to internal validity (maturation/selection by maturation, mortality, imitation, compensation, compensatory rivalry, demoralisation, regression to the mean) were not relevant to the present as they refer to methods where participants are involved over a longer period of time rather than single session, or where participants are aware of the tasks experienced by other groups. Four threats were applicable and controlled for; biases of selection; history; instrumentation; and testing. To control for selection bias, participants were randomly allocated to one the conditions in each study either a random number generation process was used to produce a non-repeating sequence which allocated participants to their condition (Study 4) or by a system-generated random routing (Study 5). History bias refers to experiences that the participants may have had before their participation that could influence their responses in a systematic fashion. In the case of this UVE study, a relevant historical concern would have been prior participation in one of the earlier studies. Participants were asked at the recruitment stage whether they had taken part in any of the earlier studies, and excluded if they had. In Study 4, instrumentation bias was controlled for by using the same equipment for each test to ensure that the on-screen presentation was the same for each participant and the hardware was identical for every trial. Finally, testing bias or the

effects of practice and fatigue were controlled by the randomisation of the order of trials within the experiments to ensure that they were counterbalanced across the whole participant population.

It is therefore felt that this application of the experimental method was appropriate and well controlled, and the results can be used to draw conclusions as to the nature of participant responses to faces varying in human-likeness.

2.3: Methodological Issues Pertaining to the Uncanny Valley

2.3.1: The Circularity Problem

An empirical test of whether the response patterns described in Mori's (1970) thought experiment actually do occur in a controlled setting should be at the heart of all research into the UVE. Is it possible to demonstrate the presence of that particular non-linear relationship between levels of human-likeness and emotional response? If the relationship can be demonstrated then it becomes meaningful to look for possible causes for its occurrence and to consider the characteristics of the entities at the crucial almost-but-not-quite human point. If no relationship occurs, then the idea of the UVE cannot be supported empirically and alternative explanations for the 'gut instinct' reaction to NHAs should be sought. At face value it seems as though it should be a trivial matter to test the principles of the thought experiment by collecting examples of different NHAs and asking people how eerie they appear. This formed a core of early research in this area. However, the present research will argue that there is a serious flaw in this approach as there is a risk of a circular argument in establishing this basic principle in that certain NHAs are judged as eerie because they fall into the UV (i.e. they appear clearly near-human rather than human or artificial), and they are considered as examples of 'things that must belong to the UV' because they look eerie. The circular argument means it is impossible to move

beyond that self-perpetuating tautology and does not aid the understanding of those NHAs or the response they provoke.

2.3.2: Operationalising the UV graph

It could be argued that there is bound to be a circularity risk in attempting to replicate a relationship sketched as a graph where the scales are unlabelled and unquantified and the points are illustrated by examples with no indication of the magnitude of the difference between them. For example, a zombie is plotted as less familiar than a corpse, but without a labelled scale, how much less familiar is it? And why would negative familiarity necessarily result in a sense of eeriness? Empirically testing the UVE calls for aspects of the original graph to be defined in more detail and this research argues that there are three components that need to be operationalised when attempting to design a study testing whether the UVE exists. Considering these will avoid the circularity problem and allow confidence that any findings are a valid and fair test of the UVE. Firstly, the human-likeness of the stimuli should be quantified either when they are generated or by independent ratings. It is important to not just to know that an agent is near-human but to be able to say empirically how human-like it is. Secondly, examples of human and non-human stimuli should be evaluated along with any NHAs. This would act as an anchoring measure to ensure that any evaluations are making use of a full scale of human-likeness rather than risking ratings grouping towards the human end of the scale. This could give misleading results and potentially magnify very small differences. Finally, consideration should be given to the emotional response that is measured when human-likeness is varied. It may be easy to find a 'valley'-type pattern as a response to varying human-likeness but to truly explore whether this is an uncanny valley, the emotional dimension needs to be explored. These considerations will now be discussed in detail.

2.3.2.1 Methods of systematically varying human-likeness and ensuring valid anchoring.

Human-likeness can be systematically varied in several ways but two will be considered here, morphing techniques and image manipulation. Firstly, in terms of morphing, identifying or generating stimuli that can be used in studying the UVE is a challenge in this area. The approach of selecting a bank of images or videos of existing androids, zombies, robots or computer game characters has been used in several studies (e.g. Ho et al, 2008; MacDorman, 2006; Tinwell et al, 2009) but given the twin issues of circularity and subjectivity this may be a risky strategy. Methods for systematically varying the human-likeness of an agent are preferable as it would always be possible to be sure where on the human-likeness axis each example could be located. One method of achieving this is to use computer software to ‘morph’ between an image of a human and a non-human face in a series of gradual steps. While it is possible to morph between a whole-body image, faces are the area of interest for the present study. The features of the starting face are gradually blended into the target face until all the original features have been replaced by the final image. It is possible to extract images at defined points during the blending process and these would consistently give a stimulus that was X% human / Y% non-human. In this way, it is possible to say with confidence that each stimulus occupies a certain place on the X-axis of the UVE graph which would then allow the plotting of a response variable against that human-likeness value. This method also ensures that the set of stimuli for any human / non-human pair will always have an anchor in their start and end images which can then be included in the experiment to ensure that comparisons are validly drawn across the whole of the human-likeness scale.

Recent technological advances in computer graphics technology have made morphing a fast and easy process to use to create stimuli . However, the initial choice of matched

human to non-human image is key to the success of any pair of images. It is important that the human and non-human images are carefully chosen and well matched for aspects such as skin tone, gaze, face shape and rough shapes/sizes of internal features. This is to ensure that any sudden and jarring transitions during the blend are avoided as the morphing algorithms can cope well with small changes but making large transitions between faces posing different expressions or looking in different directions can cause the introduction of 'artefacts' - shadows or lines on the face that did not appear in either starting or end image but which are created in an attempt to elide from one to the other. Morphing has been a popular technique for explorations of the UVE and has been used by Hanson et al, (2005) Green et al, (2008) and Cheetham et al (2013) and with carefully chosen start and end images it can be a valuable method for systematically creating near-human stimuli that also have a relevant human and non-human anchor for comparison.

Secondly, image manipulation will be considered. While morphing offers fine control of the dimension of human-likeness, there are types of stimuli that need to be created which require different techniques. In Phase 3 of this research, it was necessary to create modified human faces which blended different expressions within a single face. For example, a smiling and happy face was modified so that the expression shown in the eye region was actually angry, sad, fearful or disgusted. This was achieved using a graphics package to carefully isolate the eye region in the second face and paste it onto the first face. The edges of the pasted eye area were smoothed into the base face to ensure that the blended face looked realistic. These faces were used to test hypotheses about perceptions of emotional expression and the role of mismatched expressions in triggering the UVE. Image manipulation techniques have been used within the UVE literature previously as Seyama and Nagayama (2009) used manipulation of individual features to create faces where the eyes were exaggerated in size. Tinwell et al (2010) manipulated

video footage of facial expressions to create modified faces which did not display any expression in the upper part of the face. In both instances, and with the present research, the manipulation allowed the creation of ‘impossible’ faces that could not be experienced in nature and so were useful for isolating and testing specific aspects of faces that may contribute to the UVE.

2.3.2.2 Issues of Measurement.

The question of what should be measured and which emotions characterise the UVE is key to a valid exploration of the effect. Ziemke and Lindblom (2006) posed an initial broad question on this topic by asking whether it was possible to compare a response to an android to a response to a human and wondered what would constitute a just noticeable difference between the two. There was a particular challenge to overcome as it may be that the mechanisms which dictate how we respond to androids may not even be available to conscious introspection. Mauss and Robinson’s (2009) meta-analysis review of studies measuring emotion concluded that it was valid to ask for self-reports of emotion, especially if this was in direct response to a current stimuli. Recall of past emotions was seen as being more problematic. Mauss and Robinson also questioned whether it was most productive to conceptualise emotions as dimensions or discrete states for the purposes of measurement. Dimensions of emotion aggregate measures of the levels of several underlying factors which Mauss and Robinson suggested could include valence, arousal and motivational state. The emotional state of fear would therefore be a negative, high-arousal state with an avoidance motivation. The discrete state theory does not see emotions as combinations of points on different continua but defines them as unique entities with their own specific behavioural, physiological and subjective properties. Through their review, Mauss and Robinson concluded that

dimensions of emotion had better explanatory power for measuring emotions and their recommendation was that they represented a better approach than discrete emotions. It is therefore valid to take a scale-based approach to the measurement of emotions and this was implemented in the present research where participants were asked not only which emotion they were experiencing when looking at the NHFs, but how strongly they experienced it at that moment in time.

These considerations of emotional measurement should be put in the context of the labels and interpretation of the original UVE graph where the Y-axis was simply labelled as displaying positive and negative 'familiarity', the English label for the original Japanese *shinwa-kan*. There has been considerable discussion in the literature as to how this can best be interpreted. Bartneck et al (2007) reviewed the original translation and suggested that a better term for a Western readership would have been 'liveability' or how comfortable people would feel living with the agent in question and this was a term that informed the development of the imagined encounter scenarios in the first phase of this research. The most recent translation of Mori's original article (2012) suggests that 'affinity' may be the most accurate word to describe the intended scale and negative affinity is more easily understood as having unsettling or eerie components than negative familiarity.

Ho et al (2008) explored the emotions that were associated with negative perceptions of near-human robots. Participants viewed videos of robots (plus a human anchor) and indicated the extent to which they agreed with statements of the nature 'The figure makes me feel...', completed with one of 27 emotions. There were four groups of emotion terms: negative terms for how the figure may be perceived, (e.g. 'nauseated' and 'resentful'); positive experiential terms, (e.g. 'happy' and 'attracted'); negative terms for how the figure made participants feel (e.g. 'sad', 'loneliness'); and neutral terms

relating to surprise or confusion. As well as those general emotions, (Ho et al., 2008) also measured the UVE-specific terms of eeriness, creepiness, strangeness and human-likeness using statements of the form 'the figure looks ...'. The agreement with the emotion statements were analysed along with the UVE attributes and it was concluded that eeriness and creepiness were the best terms in describing how participants responded to the 'uncanny' robots. They found that there was a strong association between 'eerie', 'creepy' and 'fear', as well as 'shocked', 'disgusted', and 'nervous' and a weaker association with the term 'strange'. This provides support for the idea that the familiarity-strangeness measurement derived from the original graph is only capturing part of the response to the UVE and that it is always important to evaluate eeriness alongside any emotion terms that are explored. Ho et al's (2008) findings were incorporated into the analysis of the exploratory phase of the present project and provide a useful framework for evaluating those responses.

The current research presented in this thesis has been designed as a systematic exploration of NHAs created using systematic techniques to represent carefully controlled points on the dimension of human-likeness. It measures emotion and eeriness and as such explores not only whether there is a valley in emotional responses but also the extent to which it can be considered uncanny. In this way, it represents an empirical test of the UVE, uses qualitative and quantitative responses to explore the emotions that typify a response to NHAs and also explores the characteristics of those NHAs which participants found to be unsettling and eerie.

2.3.3: Near-humans as research stimuli

The final methodological aspect relating specifically to studies of the UVE is the issue of near-humans as research stimuli, and specifically the use of static pictures of NHFs.

This section will consider two issues, the use of static images over videos or embodied examples and the novelty of stimulus images within a study.

2.3.3.1 Photographs, video and embodied stimuli

The present research uses static images as the stimuli for all studies. Animation and video were considered as alternatives as they have a precedent in previous studies (e.g. Pollick (2001), MacDorman (2006), Chaminade et al (2007), Tinwell et al (2013)) and there is a theoretical argument based on Mori's original (1970) graph that motion exacerbates the steepness of the UV curve which could only be validated using moving stimuli rather than still images. Embodied stimuli were also considered and Gee et al (2005) suggest that studies examining reactions to prosthetics (limbs or partial limbs) would be of interest as embodied examples of stimuli which share many characteristics with their real counterpart, but are still tangibly different. Some case study research into the UV has been carried out with actual androids, such as Ishiguro (2005), and Becker-Asano et al (2010) but these were small scale explorations and hard to generalise. These embodied stimuli do add more information to the perceived object used in research studies but it is this complexity which might actually present a problem for analysis as it would have been harder to isolate individual features for scrutiny and draw direct conclusions from the findings. Ziemke and Lindblom (2006) make a strong argument that, in evaluating a whole embodied android, you are in effect evaluating a large number of subsystems of that android rather than any one single component so if just one of those was the component responsible for the near-human sense of eeriness, not only would it be difficult to isolate which was responsible but it might lead to erroneously concluding that the android as a whole was causing eeriness when only one sub-system had those qualities. The decision to use static images over video or working with embodied agents was made for reasons

of parsimony as well as empirical clarity as it is more complex to produce or manipulate video or to access androids than it is to work with still images.

2.3.3.2 Novelty of stimuli

A final issue of design with regard to near-humans as research stimuli relates to how images are presented to participants during a research trial and considers whether participants should view each image once or repeatedly during their participation in the study. Some of the more recent studies (Creed and Beale, 2008; Tinwell et al, 2013) presented participants with the same stimuli more than once during data collection. Using multiple presentations of the same face also poses a problem in evaluating ‘familiarity’, and indeed familiarity could easily impact on an emotional response to a face. In addition, avoiding repeated exposure to the same facial stimuli is a key element of psychological studies of face perception, other than where a repeated exposure is the subject of the research. The present research has, therefore, avoided this in the experimental studies to ensure that participants are not making relative judgments between the different presentations of the faces and instead are rating each one independently based on its individual attributes.

2.4 Ethical Considerations

2.4.1: Ethical Clearance

The studies reported in this thesis were approved by the Open University’s Human Research Ethics Committee, under references #592 and #878. The studies were carried out using the British Psychological Society’s Code of Ethics for Human Research as a guiding framework (Reeves, 2011) and were judged as presenting a very low risk to participants, especially during the web-based studies where the experiment was taking

place at a distance and there was no direct presence influencing participation. This section will detail the ethical issues that were considered in designing the present research, both in terms of general issues relevant to all psychological research with human participants and some specific points for consideration due to the nature of the UV and the materials that are used in its exploration.

2.4.2: General Considerations

The studies in this research thesis were designed to meet current high standards for ethical good practice. This section will detail the measures that were taken to ensure the ethical treatment of participants from recruitment through to participation and debriefing.

Firstly, all participants were recruited as volunteers. There was no compulsion for any individual to take part in any of the studies and similarly, no interested volunteers were denied the opportunity to take part. Participants were informed at the recruitment stage that all research data were to be held confidentially and securely. Confidentiality was maintained as no uniquely identifying information was stored with the experimental response data, ensuring that it would not be possible to identify a respondent from the answers they provided to the research instruments. The web-based studies did include the facility to leave an email address to join a mailing list and receive a report of the final results of each study but these were separated from the responses to the survey after the results had been downloaded. Data security was managed by storing all research materials on a secure server which was encrypted using a strong password and could only be accessed by the researcher.

Returning to the recruitment of participants, a detailed introduction to each study was provided to ensure that the broad purpose, scope and plans for using participant

data were clear and to provide participants with the relevant information they needed to be able to make an informed decision to take part. The details of the specific research aims and subtleties of how participants were routed through each study were not described in the introduction but were included in the debriefing at the end. Participants were advised that their participation remained voluntary all through the progress of the study and they were able to withdraw at any stage and have any data they had provided up to that point deleted by contacting the researcher. No participants chose to do this. Withdrawal was possible in the web studies simply by not completing all of the questions and there was also no penalty for not completing the questions or for simply clicking through them to the end.

Secondly, it was important to ensure that participants had a clear route for contacting the researcher with queries before, during and after the studies. With the web-based studies participants were able to make contact by email, telephone or Internet telephony from the initial launch of the study all the way through to the eventual end of the overall data collection phase. Support for the face-to-face study included conversations prior to participation as well as support by email after the sessions had taken place. This support was in place in case of questions about data handling, confidentiality and data usage as well as any concerns that the participants had as a result of their experience working with the stimulus materials. No adverse reactions or events were recorded and the contact strategy had an unexpected positive outcome in that some participants continued to make use of the contact channels to share anecdotes or experiences relating to the UVE several years after they had been involved in the studies.

Thirdly, each study closed with a detailed debriefing step which explained the research aims in the context of the UVE and detailed how the responses provided by each participant would be used in aiding understanding of this phenomenon. Where

participants had been randomly allocated to a particular experimental or control group this was explained. Participants were also given the opportunity to keep in touch for final dissemination of the eventual results. In this way, participants' last experience of the experimental setting was a positive one, hopefully leaving them with a good impression of this project in particular and psychological research in general.

2.4.3: UVE-Specific Considerations

The considerations above are standard to general psychological research and were built into the research studies as implementations of general good practice. However, the specific nature of this research project meant that there were some areas where special consideration needed to be given to how to handle potentially challenging ethical issues. These will now be described and the steps taken to address them detailed.

Firstly, any exploration of the UVE has the risk that at least some of the stimulus materials must have the potential to unsettle or disturb. If there was no possibility of the images seeming uncanny then they would not be valid for an exploration of the effect. The UVE is a subtle effect rather than one causing overwhelming fear or horror so the materials would never be shocking or terrifying, but some of the faces were certainly subtly disturbing and had the potential to elicit anxiety and nervousness in participants. This is an issue which will be faced by any research into this area and the general approach has been to ensure that although this small risk exists, participants should not be exposed to anything more distressing than they might encounter in day to day activity involving NHAs. Stimuli should be no more distressing than witnessing a computer game character, android or CG avatar. This rule of thumb combined with a support strategy to ensure that participants had a means of making contact with the researcher in the event

of any distress was deemed to be an acceptable risk management strategy when applying for ethical approval.

Secondly, while participants were not deceived as to the overall direction of the research, the specific details of how the stimuli they had encountered had been manipulated to vary their human-likeness and how this aided research into the UVE were not revealed until after they had taken part. This was to avoid any chance that participants might skew their answers to either support or contradict the UV theory if they were aware of the specific purpose. This is not an issue unique to the UVE and it is generally felt that as long as participants are debriefed as soon as possible after their participation and are given a clear route for contacting the researcher if they have any further queries then this withholding of the full detail of the purpose of the study is acceptable practice. In the present study, debriefing took place either as soon as the last question had been answered for the online studies or when the task had been completed for the face-to-face experiment.

Thirdly, the face-to-face study in Phase 2 should be considered for the presence of demand characteristics as in this case, the researcher was present while the study was being carried out. This carries the risk that participants may feel pressured to perform 'correctly', or more quickly than they would otherwise feel comfortable doing if they were not being observed. This was managed in those sessions by reassuring participants that while they were encouraged to respond as quickly as they felt able to do, they were not being personally judged on the speed or accuracy of their responses and that their individual times would be anonymised before being compared with any other responses.

In conclusion, it is felt that the methods taken to address these considerations were effective and that this research was carried out in an ethically sound manner, ensuring that no harm came to any participants as a result of their contributions.

Chapter 3: Phase 1: Exploring The Valley

3.1: Introduction to Phase 1

The first phase of research for this thesis was set in the context immediately following the Views of the Uncanny Valley workshop and Interaction Studies special edition and develops Pollick's (2005) conclusion that at that stage the UV was a descriptive rather than predictive term. To move beyond this and investigate the valley fully, it was suggested that both axes of the graph needed to be better understood. Pollick suggested that attention should be focused particularly on the dimension of 'familiarity' as the nature of the emotional response in the UVE required greater definition. The position at the close of the special edition appeared to be that the UVE had been established as a novel area for scientific enquiry with considerable interest in detailing the nature of the effect itself and establishing what was responsible for causing it. While the eerie, unsettling, and aversive nature of the UV had been described in colourful terms, there was no detailed mapping of that emotional component, or a consideration of whether it was a novel elicitation of familiar components from fear or disgust, or a specific emotion only elicited by NHAs. Further research was clearly required and several approaches had been suggested for how these enquiries could be carried out but this thesis will argue that empirical evidence at the time was limited with none of the studies published at that time adequately defining both human-likeness and emotional response and as a result, there was a clear gap in the understanding of both the emotions that were involved in the UVE and the properties of the agents that could cause it to occur.

Mori's (1970) proposal suggested a specific pattern of emotional reactions when encountering agents increasing in human-likeness. His graph presented the path that he believed this pattern would describe and provided examples of agents that he saw as

illustrative of certain points along the path. The UV concept only holds true if the conjectured graph can be reproduced from a real-world dataset by plotting the positions of the illustrative agents using data points gathered from collecting measurements of certain key variables. A study resulting in such a dataset could be designed by identifying a quality or a behaviour of a NHA that appears to contribute to the UVE. That specific quality or behaviour would be the object of the investigation but to ensure rigour in the translation process of mapping the idealised UV graph in real-world data, certain criteria would need to be set for how data were collected about those qualities.

MacDorman (2006) reported a study where ratings were collected for several video clips of NHAs to explore the relationship between human-likeness, strangeness and eeriness. While this study did not set out to offer an explanation for the UVE, it started to set out parameters for considering how to collect data about the UVE as this was the first study to propose specific scales for measuring those qualities.

A critical reading of the literature that had attempted to comment on or explain the UVE concluded that no studies were ably capturing all of the elements of the complex problem inherent in researching the UVE. The 'common sense' idea that near human-likeness would inevitably provoke an eerie response appeared to dominate the design of the research studies. It was felt that this was only able to represent part of the overall relationship between perceptions of human-likeness, eeriness, and some measured emotional response. To address this gap, the items in Table 1 below are proposed as principles for any future studies attempting to draw conclusions about the nature or cause of the UVE.

- 1) A range of stimuli should be examined, as it is not possible to draw conclusions about a continuum of human-likeness when only two illustrative points are being compared or when the continuum does not include a human anchor or control.
- 2) To emulate the graph's X axis, the human-likeness of the stimuli displaying the identified quality should be controlled or measurable. Control would involve the selection of stimuli that vary in their human-likeness in a systematic way (e.g. CG agents or morphs between human and non-human agents) but if this is not possible, the stimuli should be independently rated to measure their subjective human-likeness. Without a measure of the human-likeness of the stimuli, a study risks being an arbitrary inventory of the qualities of NHAs.
- 3) The Y axis of the graph, labelled as 'familiarity', represents an emotional response or reaction. This is not as clear-cut as the human-likeness dimension as there are many different interpretations of what could be meant by this axis. Therefore, an effective study of the UVE should clearly identify the emotion that is being measured for this aspect. To ensure psychological validity, response scales that are used to measure this aspect should be grounded in established psychological evidence relating to human emotions.
- 4) When the two measurements of human-likeness and emotional response are plotted against each other, the path described by the response measurement should display a clear dip or deviation from a linear path, occurring between 50% and 100% on the human-likeness axis.
- 5) Alongside the measurements of human-likeness and emotional reaction, a specific rating of eeriness should also be collected for each of the stimuli. There are two reasons for this. Firstly, without it, any observed valley could be explained as a mere dip in appreciation or liking and it would not be possible to confidently assert that the valley was uncanny in its nature. Secondly, it avoids the risk of a circular approach as described in Chapter 2 above as by including a direct measurement of eeriness, the researcher can be more confident in the conclusions that are being drawn about the qualities under examination.

Table 1: Research principles for investigating the UVE

If, when these factors are controlled and measured, a pattern can be plotted which approximates the path described by Mori, with the stimuli falling into the valley also displaying high ratings of eeriness, then an UVE can be said to have been found and conclusions can be drawn about the qualities of the stimuli which have elicited it. Studies which do not display this careful control may be useful in collecting other forms of evidence about NHAs but cannot be said to contribute to an understanding of the UVE *per se*.

Considering the studies discussed in Section 1.1: Locating The Uncanny Valley Effect, it can be seen that the empirical evidence was limited with none of the reported studies

meeting the criteria set by the design principles above. None adequately controlled for both human-likeness and emotional reaction with concurrent measurements of eeriness and the selection of stimuli was small scale and in some cases appeared arbitrary. The studies can be considered in three categories: those that offered a theoretical perspective on the UV but did not report primary data collection (Brenton et al, 2005; Gee, 2005; Ramey, 2005; Ziemke and Lindblom, 2006); those that applied existing human-likeness or anthropomorphic research to the UV concept (Chaminade et al, 2005; Keyzers and Gazzola, 2005; Pollick 2005) and those that report primary data collection about NHAs. Of those, Berthouze and Berthouze (2005) collected data about the qualities of agents that varied in human-likeness but did not conduct any measurement of emotion or eeriness and did not demonstrate the presence of an uncanny curve. Hanson et al's (2005) second study uses a morphing technique to standardise the human-likeness dimension but then only measured acceptability so no conclusions of uncanniness can be drawn. In addition, the study in question used a single continuum so the observed ratings of acceptability can only be applied to that one instance of human to cartoon morph. Finally, both Ishiguro (2005) and MacDorman (2005c) reported data collection exercises that did not employ a range of stimuli varying in human-likeness as Ishiguro (2005) used a single example of an android, and MacDorman (2005c) a single android paired with a human anchor.

Interest in this area was also informed by research carried out by the author in fulfilment of an earlier qualification (Gray (2006), [unpublished MSc dissertation]). The aim of that study was to take an empirical approach to the concept of the UVE using morphed images of faces, where artificial faces were gradually transformed into human faces and static images were taken at set points within that process. These produced a set of images which each represented a defined level of human-likeness, and were used in an online experiment testing whether it was possible to identify a point where participant's

emotional responses changed from a general increase in positivity to indicating a sudden ‘uncanny’ downturn indicating a more negative response. A downturn in emotional response was found but this was placed at the midpoint between artificial and human. Reflections on the research design identified several areas for further exploration as the study had demonstrated that the non-linear relationship between human-likeness and emotional response did suggest a pattern indicative of the UVE but it was limited in its scope and did not attempt to explore the reasons why this effect may occur or explore the nature of the emotions involved beyond classifying the responses as positive or negative. Therefore, it can be seen that at the inception of this Phase in 2006, none of the available evidence provided a sufficient explanation for the UVE and as a result, the cause of the UVE was unexplained and the emotional components of the effect were not known.

The overall aim of this thesis is to advance understanding in these two areas, and this first research phase established the first steps towards addressing this gap. In Study 1 (Section 3.2: Study 1: Subjective ratings of near-human agents), a collection of varied NHFs were rated for subjective measurements of human-likeness, strangeness and eeriness, using the scales established by MacDorman (2006) which satisfied the design criteria 1) 2) and 4) from Table 1 above. Conclusions drawn from these measurements led to Study 2 (Section 3.3: Study 2: Imagined Encounters with near-human agents) where three agents were chosen for further exploration along with a non-human and human anchor for comparison. This study was designed to contribute to an understanding of the Y axis of Mori’s chart (criterion 3) by detailing how participants reacted emotionally to each of the agents and introduced a consideration of the nature of NHFs by collecting data on how participants described each face. By analysing measurements it was possible to overlay real-world response data on to Mori’s hypothesised UV curve, thus allowing

conclusions to be drawn about the qualities of the chosen stimuli and satisfying criterion 5). In this way the first research phase began to explore the nature of the emotional reaction involved in the UVE and provide a basis for further studies to elucidate its causes.

3.2: Study 1: Subjective ratings of near-human agents

3.2.1: Introduction to Subjective ratings of near-human agents

The first study in this phase carried out a survey of the eeriness, human-likeness and strangeness of a collection of NHAs, using the principles outlined in the Introduction above and adopting the scales used in MacDorman (2006). As with MacDorman et al's study, this first study was not designed to provide an explanation of the UVE, but rather to begin to collect evidence for later studies which would fulfil that function. The core aim of this study was to satisfy design principles 1) 2) and 4), and thus establish a baseline of ratings for stimuli images to be used in future studies exploring the nature of emotional responses to 'uncanny' agents compared to their non-uncanny counterparts. This introduction will describe the measurement scales that were used and outline the research question and hypotheses that were addressed by this study.

Measurement scales

Three scales that had been demonstrated as effective in MacDorman (2006) were adopted for this study. In the original study, they were used to measure the human-likeness, strangeness and eeriness of several video clips of robots. The aim of MacDorman et al's study was to test whether the pattern of those responses would match the UV proposal by showing the type of negative peak that Mori had described. The three scales were an interpretation of Mori's description. Firstly, it was necessary to measure human-likeness as Mori observed that the effect occurred as entities become less mechanical

and more human-like so MacDorman anchored the scale as 'very mechanical' at point 1 with 'very human' at point 9. The other two measurements (strangeness and eeriness) were MacDorman's means of operationalising the y-axis on that chart. The strangeness scale was labeled based on Mori's use of the term 'familiarity' to describe how people felt about an entity as human-likeness changes. Mori's graph showed positive and negative familiarity on this axis, but MacDorman believed that 'strangeness' would be a more intuitive factor for participants to rate than 'negative familiarity'. The eeriness scale allowed MacDorman to test the relationship between a measurement of strangeness and the particularly unsettling quality that the uncanny valley seemed to elicit as something measured as strange alone is not necessarily eerie. Both Y-axis scales ran from 1-10. MacDorman's study provided evidence that these measurements had been successful and reliable in examining responses to comparable materials and so they were adopted for the present study.

Research question

What will the subjective ratings of human-likeness, strangeness and eeriness be for an identified collection of NHAs?

Hypotheses

H1. This hypothesis tests whether two measurements of emotional response (strangeness and eeriness) vary with measurements of human-likeness and tests whether such a response could be considered uncanny.

When considering the ratings for each agent, the following overall patterns of correlation will be found:

- Human-likeness and strangeness will be strongly negatively correlated as this hypothesis is based on the overall shape of Mori's graph which requires that as human-likeness increases, familiarity should also increase. MacDorman's (2006)

scale uses strangeness as the antonym to familiarity in measuring this axis, so as familiarity increases, strangeness should decrease.

- Eeriness and strangeness will be strongly positively correlated as Mori's graph is only labelled according to familiarity but his description of the concept evokes ideas of eeriness. To be able to accurately describe a response to changes in human-likeness as 'uncanny', it should follow that ratings of eeriness should increase as ratings of strangeness increase.
- Human-likeness and eeriness will be strongly negatively correlated as if there is a 'valley' effect, and it can be understood to be uncanny, it would need to follow that as human-likeness increases, eeriness should decrease.

H2. This hypothesis is testing for whether a 'valley' pattern can be seen to emerge in the relationships between human-likeness and strangeness or eeriness. When the mean ratings for each agent are plotted against each other, the strangeness and eeriness ratings will show a 'dip' or valley in their linear trend and this will occur somewhere between the 50% and 100% on the human-likeness scale.

3.2.2: Method

Design

This study used a repeated measures design to collecting ratings of 20 NHFs on three dependent variables: human-likeness, strangeness and eeriness.

Materials

Twenty images were selected from a collection of 'almost human' faces. Several types of NHA were used, and included androids, CG artwork and digitally manipulated faces.

The rating scales were taken from MacDorman (2006):

Human-likeness: Very mechanical (1) to very human-like (9)

Strangeness: Very familiar (1) to very strange (10)

Eeriness: Not eerie (1) to very eerie (10).

A Microsoft Word document was produced which presented standardized instructions, followed by all 20 images accompanied by the three rating scales.

Participants

Eight participants judged all 20 images. They were acquaintances of the researcher and 7 were employed at a UK university. All had an interest in psychological research, but none had studied psychology, or had specific or expert knowledge of the uncanny valley. They were asked to volunteer their time, and received no reward for their participation.

Procedure

Once they had agreed to take part, participants were sent the Word document by email and asked to work through the document, rating each image on each of the three scales. They also had the opportunity to add comments about each image if they wished. The document was then returned by email.

3.2.3: Results

This study was an initial step towards answering the first research question in Phase 1: What will the subjective ratings of human-likeness, strangeness and eeriness be for an identified collection of NHAs? Table 2 presents the overall means and standard deviations for the measures, and Table 3 presents the mean values on each measure for each image.

| | N | Mean | Deviation |
|----------------|-----|------|-----------|
| Human-likeness | 160 | 5.4 | 2.8 |
| Strangeness | 160 | 5.2 | 2.8 |
| Eeriness | 160 | 5.1 | 2.9 |

Table 2: Overall means and standard deviations for ratings of human-likeness, strangeness and eeriness






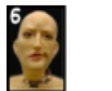
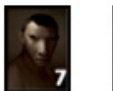
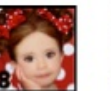
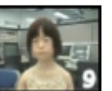


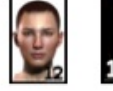








| Picture | 1 | 2 | 3 | 4 | 5 |
|----------------|---|---|---|---|--|
| Human-likeness | 4.5 | 8.4 | 5.3 | 4.8 | 1.4 |
| Strangeness | 5.1 | 3.9 | 5.1 | 5.6 | 8.1 |
| Eeriness | 6.9 | 2.8 | 4.9 | 4.9 | 6.9 |
| Illustration |  |  |  |  |  |
| Picture | 6 | 7 | 8 | 9 | 10 |
| Human-likeness | 2.1 | 5.0 | 5.5 | 5.5 | 5.6 |
| Strangeness | 5.9 | 7.8 | 4.8 | 6.6 | 4.9 |
| Eeriness | 5.6 | 8.1 | 5.1 | 6.4 | 5.4 |
| Illustration |  |  |  |  |  |
| Picture | 11 | 12 | 13 | 14 | 15 |
| Human-likeness | 4.9 | 7.6 | 7.6 | 1.3 | 5.4 |
| Strangeness | 5.1 | 4.1 | 3.8 | 8.6 | 5.9 |
| Eeriness | 6.1 | 4.0 | 4.0 | 7.1 | 5.3 |
| Illustration |  |  |  |  |  |
| Picture | 16 | 17 | 18 | 19 | 20 |
| Human-likeness | 2.0 | 5.4 | 8.8 | 8.4 | 8.9 |
| Strangeness | 8.3 | 5.6 | 1.4 | 2.6 | 1.5 |
| Eeriness | 8.6 | 5.6 | 1.6 | 2.0 | 1.4 |
| Illustration |  |  |  |  |  |

Table 3: Mean human-likeness, strangeness and eeriness by image

When analysed using Pearson's r correlation, significant relationships were found between all three measures and these did indeed follow the hypothesised patterns. There were negative correlations between human-likeness and strangeness ($r = -.681$,

$N=160$, $p < 0.005$, one tailed) and human-likeness and eeriness, ($r = -.615$, $N=160$, $p < 0.005$, one tailed). A positive correlation was found between strangeness and eeriness: $r = .777$, $N=160$, $p < 0.005$, one tailed.

A second key area of interest here is in the relationship between human-likeness and strangeness, human-likeness and eeriness and the extent to which this supports the concept of an UVE. The two plots below visualise those relationships. H2 predicted that *when the mean human-likeness and strangeness ratings for each agent are plotted against each other, the strangeness and eeriness plots will show a deviation from a linear trend between the 50% and 100% human point.*

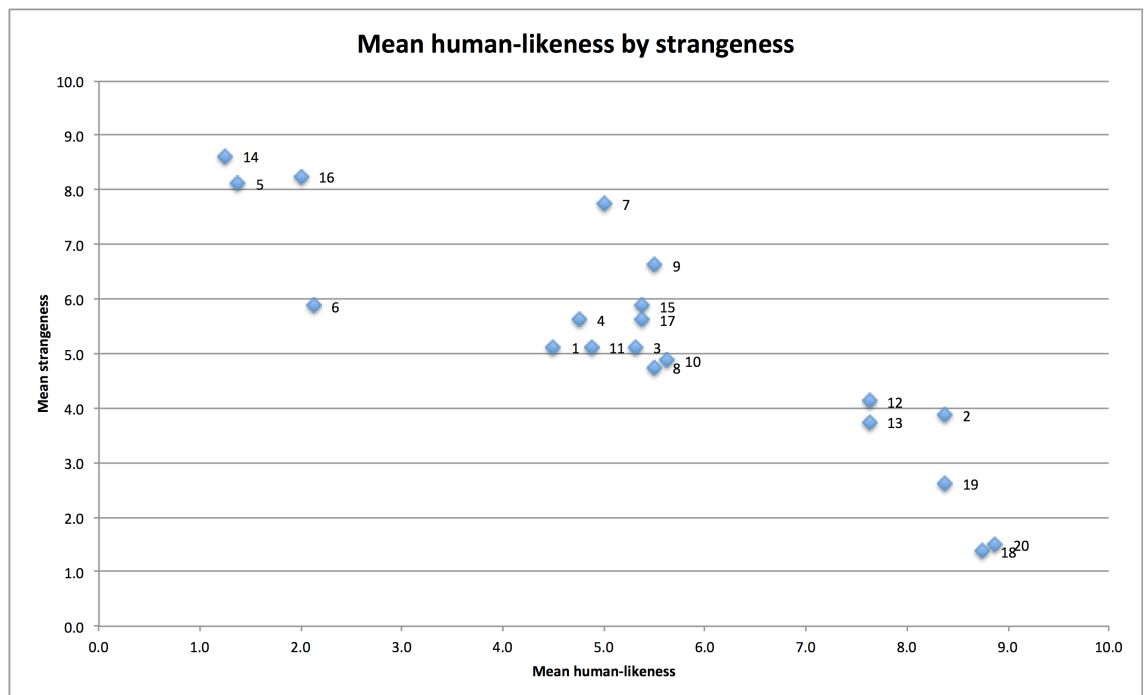


Figure 5: Scatter plot of mean human-likeness and strangeness: data labels refer to picture references.

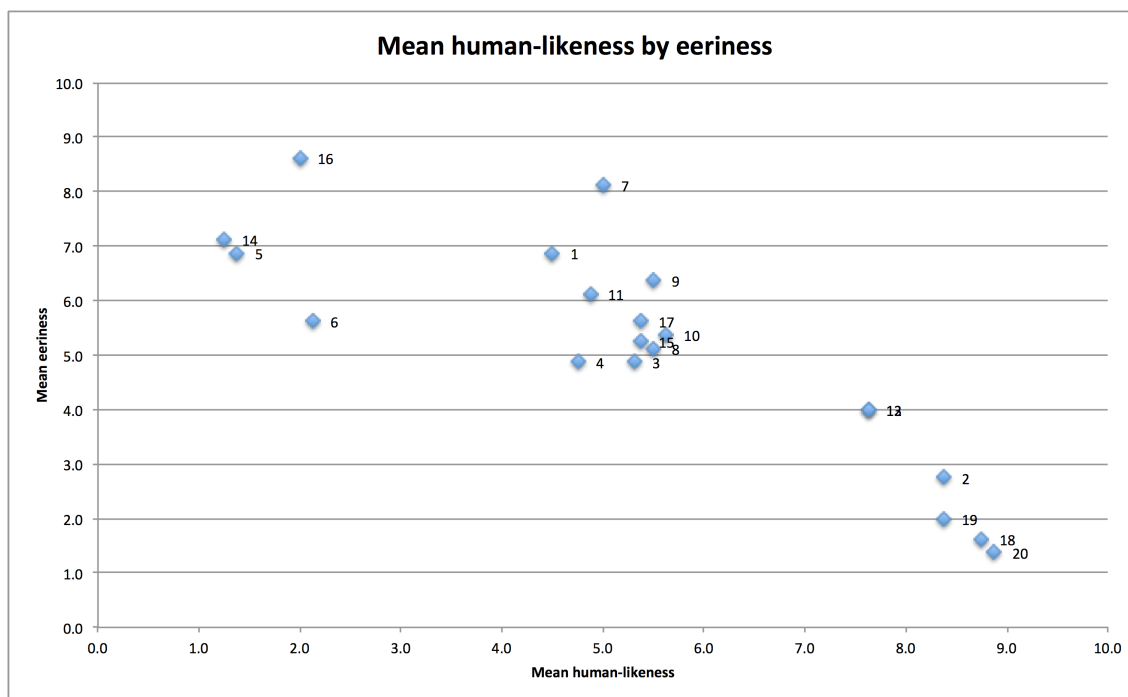


Figure 6: Scatter plot of mean human-likeness and eeriness: data labels refer to picture references.

As can be seen in the figures above, the mean ratings cluster around three distinct areas so that a deviation from a linear trend cannot be determined. The implications of this and of the earlier findings will now be discussed.

3.2.4: Discussion

The goal of this first study was to explore ratings of a variety of near-human images to see whether Mori's hypothesised relationship between human-likeness and familiarity could be reproduced using real-world examples and whether there was any evidence to support the anecdotal suggestion that some of those near-human images would be found particularly eerie. Subjective ratings of human-likeness, strangeness and eeriness were collected from participants who looked at 20 different near-human images collected from a range of sources including images of robots, computer game characters and dolls. The UV hypothesis suggests that two features should be found in the results. Firstly, that there should be a specific type of relationship between human-likeness, strangeness and

eeriness and secondly, that faces rated as the eeriest would fall in a region close to but not quite human.

Hypothesis 1 tested the first of these by looking at whether two measurements of strangeness and eeriness varied with measurements of human-likeness. Support was found for the hypotheses that had been proposed for all three relationships. Firstly, the UVE requires that overall, as human-likeness increases, strangeness should decrease and so a negative correlation should be found. This relationship describes the overall shape of the UV graph and is a necessary condition for a valley to emerge at some point along that continuum between non human and human. This significant negative correlation was found, and it had a moderate explanatory strength with 46% of the variance explained, providing evidence that the relationship between human-likeness and strangeness is as predicted by the UV hypothesis.

Eeriness ratings were also collected, as per MacDorman's (2006) study, in order to be able to go beyond the identification of a relationship between human-likeness and familiarity and see whether the increasing strangeness can be said to be representative of an 'uncanny' valley. This would also help to disambiguate the 'familiar-strange' axis proposed in the original graph. Without a measure of eeriness or a comparable term, even if a valley-type pattern was found, it may be that participants were interpreting familiarity in the sense of being well acquainted with a particular face image rather than interpreting it as the opposite of strange. Support was found for this hypothesis as eeriness and strangeness were indeed positively correlated. This was actually a stronger correlation than the one between human-likeness and strangeness, with 60% of the variance in the strangeness ratings explained by the ratings of eeriness. This allows two conclusions, firstly, it confirms that it is valid to talk about eeriness and strangeness as related terms in the evaluation of NHFs and secondly, that the images selected for this

study are good examples of entities which may evoke eerie responses and so are valid for an exploration of the UVE. A final correlation was performed to test the prediction that a negative correlation would also be found between human-likeness and eeriness as a more human-like face should seem less eerie according to the UV hypothesis. This was again found, explaining 38% of the variance between the variables. It can be concluded from these correlation results that the subjective ratings for human-likeness, strangeness and eeriness did conform to the first part of the UV hypothesis.

The second part of this study then went on to look at whether a 'valley' pattern can be seen to emerge in the relationships between human-likeness and strangeness or eeriness. This was tested by plotting the mean ratings for those value pairs for each of the near-human images. It had been expected that there would be a clear pattern in the plots where increasing human-likeness would lead to a decrease in strangeness, but only up to a certain point where there would be a sudden local increase in the strangeness ratings. It was expected that some of the faces would be rated as very nearly human/highly strange and very nearly human/highly eerie. However, this pattern was not found. Instead, three distinct clusters were formed of low human-likeness/high strangeness, moderate human-likeness/moderate strangeness, and high human-likeness/low strangeness, and low human-likeness/high eeriness, moderate human-likeness/moderate eeriness, and high human-likeness/low eeriness.

A goal of this first study was to provide baseline ratings of stimulus images with the intention of using them later in experiments to examine if there are significant differences in the description of and reaction to different types of near-human face. While it can be concluded that the pattern of responses found in these results does not support the location of faces that would be predicted by the UV hypothesis, the pattern of three distinct clusters is a notable finding and it is possible that within those clusters,

some of the images could still be considered uncanny. One approach to this is to review prior research which did find a UVE in similar rating data and compare the mean values found with those that were found in this study. In this way, an examination of the original UVE graph and the measurement scales as used by MacDorman (2006) can be used to develop criteria for faces that could be classified as uncanny based on the area of the chart. In this way, faces could be classed as uncanny if their mean human-likeness was between 4.5 and 8, and their mean strangeness or eeriness was over 6. In the present study, 3 faces met those criteria and are shown in Table 4 below.



Table 4: The three images (11, 7 and 1) rated as medium-high human-likeness with high strangeness or eeriness.

These all came from the middle cluster, so using this method the moderately human cluster may be considered the one most likely to contain faces which would be perceived as uncanny. The study to be described in the next phase uses an example from each of the clusters and compares rating and behavioural data to explore whether any significant difference will be found between the clusters.

A final point to note in discussing this study is that it was rather small in scale, with only a small number of participants (8) providing rating data for the faces. This study has provided an initial mapping of these three key ratings for a variety of different NHFs but later studies which will aim to explain the nature of the UVE will seek to produce more robust findings by recruiting a larger number of participants. The first of these explanatory studies will now be introduced.

3.3: Study 2: Imagined Encounters with near-human agents

3.3.1: Introduction to Imagined encounters with near-human agents

It has already been argued that at the time of the first phase of this thesis, the UVE was a descriptive rather than a predictive term. The details of the emotions involved in the effect were based on arbitrary agent selection and insight-based conclusions drawn about the agents said to trigger it. Study 2 was designed to test these assumptions by gathering evidence about the emotions involved in the UVE and exploring possible causes for its particular pattern of emotional reactions. It developed the findings of Study 1 using three of the stimuli rated in that study, plus one human and one non-human anchor, to ask detailed questions about how people felt when they imagined an encounter with those agents and the techniques they used when asked to produce descriptions of them. This work was necessary to begin to address the third design principle detailed in Section 3.1 which described that, in order to validate Mori's chart, an empirical measure of emotional response is required. By identifying the emotions associated with different NHAs, it becomes possible to derive such a measure and use it to test which qualities of NHAs are most likely to cause a UVE. This study comprised three tasks, a rating task applying the standardised scales used in Study 1 and two tasks where participants imagined an encounter with a NHA and answered questions about that encounter. This introduction will describe the background to those imagined interaction tasks and describe the stimuli that were used before detailing the research questions and hypotheses that were addressed in this study.

Imagined Interaction Scenarios

The scenario that participants were asked to imagine in this study involved an exhibition of robots and other NHAs and participants were asked to either describe each

agent in sufficient detail for a friend to be able to find them in this crowded public space, or imagine how it would feel to have this agent as a domestic robot at home. This scenario-based technique borrowed aspects of the social cognitive concept of Imagined Interactions (IIs). Edwards et al (1988) and Honeycutt (1990) suggested that IIs acted as an element of the cognitive process of communication where people ('actors' in the terminology of the theory) use a mental representation of a social encounter to rehearse for a future experience or ruminate on a past incident. Edwards et al and Honeycutt used questionnaires to explore the frequency and functions of IIs by asking closed questions on how often participants engaged in IIs as well as several free-text questions which were then coded for further analysis. They found that IIs were commonly reported and that certain themes emerged. Romantic and relational themes were common in their population but other studies have examined IIs relating to family life and employment scenarios. The researchers drew a clear differentiation between IIs and fantasies as IIs had a different and preparatory function and often included mundane activities or interactions with negative consequences which did not take place in most fantasy scenarios. However, a question remained as to whether IIs were confined solely to an imaginative realm and to explore this, Gotcher and Edwards (1990) compared IIs with actual communications and found that there was a positive correlation between the frequency and content of IIs and with conversations that actually took place, suggesting that IIs serve as practical rehearsals when an interaction that has been imagined takes place in reality. These early findings were also supported by Honeycutt (2003) who found that IIs were used on a daily basis by most people, and in a cross-cultural study by McCann and Honeycutt (2006) who found that IIs were regularly used by US, Thai and Japanese participants but that the functions varied according to the cultural background of the participants.

The evidence established IIs as a cognitive phenomenon in common use and in terms of the present study, it was felt that asking participants to self-report their IIs would be a valid way to gain insight into how they might feel about encounters with NHAs. The present study was the first to apply this type of methodology to a study of the UVE but it was felt that this was a valid approach for four reasons. Firstly, it had been demonstrated that IIs could mirror how someone would actually feel and react in a genuine encounter. Secondly, asking participants to report the content of their imagined encounter produces a rich source of themed qualitative data. Thirdly, the present study had the aim of exploring more about a topic where certain assertions had been made, but these had not yet been definitely tested. IIs are unique to the participant and as long as the scenarios describing an interaction are free from leading language, a self-report of an II represents a rich source of data about the interaction but without the participant being unduly influenced about what to say about each of the agents. Finally, on a practical level a report of an II can be collected easily from participants in their own time and at a distance and so lent themselves well to online data collection methods used in this phase.

A scenario for eliciting face descriptions and emotional responses.

This study asked participants to produce two types of description of each of the five agents. These were a description of the agent's face and a response to how they would feel if they were asked to spend time in close proximity to the agent. From a theoretical perspective, describing to and reacting to a face are explored separately in recognition of a long-running debate in the literature of face perception. The cognitive psychology of face perception has described human abilities in recognising, remembering and reacting to faces in terms of models of face processing as process flow descriptions which serve as a representation of the neural functions and information flow involved when face processing occurs. Developments in face perception theory have often been marked by

revisions to the contents or arrangement of these models. The classic model of face perception (Bruce and Young (1986)) proposed that face processing could be understood as a pathway involving several systematic steps, with information passed between each discreet module to eventually arrive at a face identification. A refinement to this model was proposed by Breen et al (2000). In order to explain delusional misidentification syndromes (such as Capgras syndrome) in which activation of a 'face recognition unit' led to two outputs, one to a 'personal identity node' and one to an 'affective response'. In normal function, the affective response helps differentiate between familiar and unfamiliar faces, but in people suffering from certain delusional misidentification syndromes, the link between the face recognition unit and affective response is damaged, meaning that although the person can recognise the face and knows it is familiar they do not get the expected emotional response.

The concept that the identity and emotional aspects of a face might be dealt with separately has been explored in research looking at whether there may be two routes or pathways involved in face processing, each responsible for the transmission of different types of information about the face stimuli. The 'dual-stream hypothesis' (originally proposed by Bauer (1986) and Ellis and Young (1990), suggests that in addition to a pathway processing information about the identity of a face, there is a second pathway which is responsible for processing the information about the emotional reaction displayed in a face. Due to the locations of the brain structures involved in the two pathways, the routes were identified as the ventral (identification) and dorsal (emotion) pathways. The regions involved are illustrated in Figure 7 below.

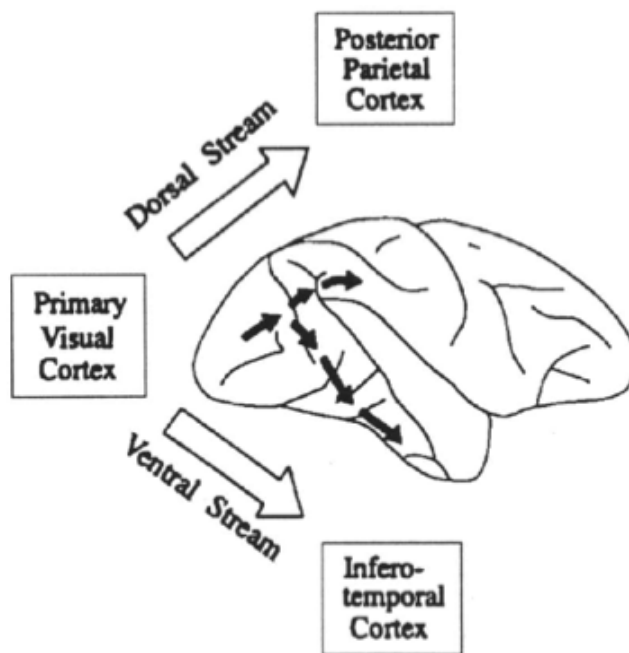


Figure 7: An extract of a diagram taken from A.D. Milner and M. Goodale (1995), *The Visual Brain In Action*, demonstrating the dorsal and ventral streams of visual processing in the macaque monkey brain. Reproduced by permission of Oxford University Press. (From Breen (2000))

The original evidence for the different pathways was drawn from studies of patients with two types of face processing disorder meaning that one of the pathways was unavailable to them. Initially, Bauer (1986) observed covert face recognition in patients diagnosed with prosopagnosia where some patients displayed a different skin conductance response (SCR) to familiar faces compared to unfamiliar faces as even though there was no conscious awareness of recognition, a different level of arousal is evoked for familiar and unfamiliar faces. This observation led Bauer to propose this dual route for face recognition with a main ventral route responsible for overt recognition and a secondary dorsal route of covert face recognition where this secondary route carried the emotional significance of a face and in the non-clinical population this additional information supports the main route for recognition. In patients with prosopagnosia where the main route is not accessible, the dorsal route alone gives rise to the non-conscious reaction described as covert recognition. Ellis and Young (1990) developed this

dual-route hypothesis further by synthesising Bauer's evidence from prosopagnosia with observations from studies of patients with Capgras' syndrome as these patients have intact face recognition ability but an abnormal response to faces familiar to them, thinking that they represent duplicates or imposters. It was later found by Ellis et al (1997) that patients with Capgras syndrome did not show a difference in SCR for familiar faces, providing support to the idea that information about the emotional significance of the face via a dorsal route is not operational for those patients. In terms of the dual hypothesis theory these patterns of preserved and lost perceptual abilities in Capgras syndrome could be said to represent a mirror of the covert recognition present in some prosopagnosic patients as in prosopagnosia, the emotion-route is intact even though the perception-route is not. In Capgras syndrome, the perception-route is preserved but the emotion-route is disrupted. Considered as a pair, those two clinical presentations mirror each other in the types of face recognition abilities preserved and lost and this pattern was used as evidence for the dual route hypothesis.

| Disorder | Prosopagnosia | Capgras Delusion |
|-------------------------|---|---|
| Intact ability | Emotional connection to familiar faces | Conscious recognition of familiar faces |
| Lost ability | Conscious recognition of familiar faces | Emotional connection to familiar faces |
| Functioning pathway | Dorsal | Ventral |
| Non-functioning pathway | Ventral | Dorsal |

Figure 8: Summary of research findings relating lost and preserved abilities to the dual pathway hypothesis.

Breen et al critically approached this argument. Firstly, they note that the evidence at the heart of this idea is the SCR which is measured for both type of patients as in itself it doesn't actually indicate recognition, it is a measure of arousal only. The magnitude of the SCR is also markedly smaller in the prosopagnosics compared to normal controls. Secondly, they question the evidence for the designation of the two routes as perceptual

versus emotional and note that the original neuropsychological evidence cited by Bauer contained more detail on the emotional processing of different systems than is present in Bauer's conception of the dual route hypothesis and that Bauer's hypothesis conflates affect and arousal. To address these perceived limitations of the dual route model, Breen et al propose a new model. They suggest that in both normal and prosopagnosic or Capgras brains, a single neuroanatomical route across multiple structures explains face recognition. The pattern of preserved and lost abilities in Capgras and prosopagnosia result from problems in passing the information between the ventral temporal lobe structures and the ventral limbic system. In contrast, they proposed a new dual-route cognitive model based on this concept with familiar and unfamiliar face recognition taking different routes.

Debate continues into the validity of the dual route hypothesis, with Young (2009) emphasising the similarity in the subjective experience in patients with each type of delusional misidentification disorders and explored what could be understood from their differing experiences of 'familiarity'. However, and in terms of the present study, presence of a lively debate into whether emotional responses to faces are processed separately to the more cognitive recognition, identification and ability to describe a face does support that it is valid to examine these two areas through separate tasks. To this end, separate tasks were designed to explore how participants would feel when presented with the different NHAs, and how they would approach being asked to provide a detailed description of each one.

Emotion task

The emotion task was informed by the common-sense assumptions used to describe the emotions that characterise the UVE and was designed to elicit free-form descriptions of emotional responses to the different agents. While negative reactions were expected

to the NHAs, the question wording was neutral so positive reactions could also be given if participants felt warmly towards them and to ensure there was no influence due to leading questions. The task asked how participants would feel about sharing their home with the NHA which was chosen to elicit deeper feelings than a simple one-off encounter and also the presence of the agent in a domestic setting was designed to evoke a sense of the extent to which participants felt they would be able to trust the NHA when present in a setting where they could expect to be comfortable.

Description task

In this task, participants were asked to provide a description of the agent, sufficiently detailed as to allow someone else to find them in a crowded arena. The purpose of this task was to see how participants would approach the task of describing the agents, with the specific research aim of whether they would concentrate more on descriptions of individual features or of an impression of the face as a whole. It has been described above that human faces are generally processed holistically rather than analytically so the results of this task will allow an examination of whether this tendency holds for NHAs. It may be that there is a division between human and NHAs where faces stop being processed as mechanical objects and become processed holistically and this task will allow an initial exploration of that relationship by looking at the frequency of references to different parts of the face.

Stimuli

Five faces were used as stimuli for this study. Three were taken from the image clusters as identified in Study 1, one from each of the barely humanoid, moderately humanoid and highly humanoid clusters. The moderately humanoid cluster contained the three faces (out of 20) which were found to meet theoretical criteria for uncanniness and so it is theorised that this cluster is the most likely to represent uncanny faces. To ensure

that this study meets the first design principle described in the Section 3.1 above, two anchors were added to the stimuli set, a robot face to represent an entirely artificial face and a natural human face to act as the entirely-human anchor.

Research Questions and Hypotheses

Rating task

RQ1: What will the subjective ratings of human-likeness, strangeness and eeriness be for an identified collection of NHAs?

H1. Face type will have a significant effect on the ratings of human-likeness, strangeness and eeriness.

H2. When considering the ratings for each agent, the following patterns of correlation will be found:

- Eeriness and strangeness will be strongly positively correlated
- Human-likeness and strangeness will be strongly negatively correlated
- Human-likeness and eeriness will be strongly negatively correlated

H3. When the mean human-likeness and strangeness ratings for each agent are plotted, the strangeness plot will show a deviation from a linear trend between the 50% and 100% human point.

(These are the same correlations and plots predicted in Study 1 above and grounded in Mori's UV graph and MacDorman's (2006) scales derived from this, so the rationale for these predictions is the same.)

Description task

RQ2: When asked to describe the faces of five NHAs, which features will be emphasised and how will this compare to the descriptions of human and non-human comparison faces?

H4. When descriptions are analysed for the frequency of types of descriptive term, there will be a significant difference between the frequency of terms describing the whole face compared to those describing individual features. The rationale for this hypothesis is to begin to explore whether there is a relationship between levels of human-likeness and how faces are processed and particularly whether there was any evidence that NHFs were processed analytically rather than holistically. At this stage, while this was an area of interest, there was no suggestion from the supporting literature relating to the UVE that human-likeness, eeriness and analytic processing, so no specific predictions were made about the nature of that difference.

H5. When the descriptions are analysed for the frequency of references to different facial features, there will be a significant difference in the number of references to the eyes by face type. This hypothesis is grounded in the idea that it may be an aspect of how eyes are presented that can make NHFs appear eerie because as the moderately human face was drawn from the cluster containing the most likely UV candidates, it is predicted that the references to eyes will be highest for this cluster.

Emotion task

RQ3: When considering a hypothetical social encounter with the NHAs, which terms will be used to express how people would feel about this and which emotions will characterise their reactions? What emotions characterise the reactions to the five faces? What pattern of negative, positive and neutral words will be used in people's reactions, and will the negative words vary in accordance with a hypothesis that the UV is characterised by 'creepiness' or 'unease'?

H6. The response descriptions will be analysed for how frequently positive, negative and neutral emotion terms are used to describe each face:

- a. The highest frequency of negative emotions will be used for the moderately human face type.
- b. The lowest frequency of positive emotions will be used for the moderately human face type.
- c. There will be no significant difference in the number of neutral emotion terms used for each face type.

The rationale for these predictions is that the moderately human face was drawn from the cluster containing the three best UV candidates. The concept of the UV requires that there is a negative or unsettling emotional component to the response to NHFs so that would predict that the moderately human face would be described more negatively and less positively.

H7. Emotion terms synonymous with eeriness or unease will be highest for the moderately human face type. This develops H6 and is a more direct check of whether the UV candidate face type was actually described as more eerie when viewed by participants.

Triangulating Across Tasks

RQ4: This question combines results across tasks to develop H4, H5 and H7 and understand the relationship between eeriness ratings for each face type and the measured references to individual features as opposed to whole face descriptors, references to eyes, and emotional responses containing words synonymous with eeriness.

H8. A significant correlation will be found between ratings of eeriness and the frequency of descriptions referring to individual features rather than to the whole face: the rationale follows from H4 but this looks at the eeriness ratings as opposed to

predicting based on the moderately human face type as the potential UV candidate. As above, no specific prediction about the direction of the correlation was made.

H9. There will be a positive correlation between ratings of eeriness and the frequency of references to eyes which develops the concept detailed in H5 above that one of the explanation for the eeriness in the UVE may be some aspect of eye appearance where higher ratings of eeriness are linked to referring to the eyes more in the description.

H10. Ratings of eeriness will be positively correlated with the proportion of descriptions containing a word synonymous with uncanniness. The rationale here is to examine whether the eeriness ratings were in line with the emotional response behaviour as it would be expected that as rated eeriness increases, so would the emotional response behaviour.

3.3.2: Method

Design

This study used a repeated measures design to explore the within participants factor of 'face type', with five levels representing an example of either a potentially 'uncanny' face or a human or artificial anchor face. The levels were identified as Artificial, Cluster One, Cluster Two, Cluster Three and Human-like: cluster one represented faces rated low for human-likeness and high for strangeness/eeriness, cluster two faces that were medium human-likeness and medium strangeness/eeriness and cluster three faces that were highly human-like but low on strangeness/eeriness. The same design was adopted for three different tasks that were completed in the same session by participants. The following dependent variables were drawn from the three tasks:

Description Task

The first dependent variable from the description task counted references to specific facial features or characteristics for each face type. The free text comments were coded and the numbers of references to gender, race, age, whole face, eyes, mouth, nose, hair, skin, chin, cheeks, ears and accessories were counted. These counts were also used to derive a second dependent variable categorising whether the references were to the whole face or to individual features.

Emotion Task

The first dependent variable from the description task counted references to different feelings that participants expressed about the different face types. The free text comments were coded and the numbers of references to seventeen different emotions were counted. These counts were also used to derive a second dependent variable categorising whether the references were positive, neutral or negative in nature.

Rating Task

The rating task contributed three dependent variables, ratings of each face type on three scales: human-likeness (H) strangeness (S) and eeriness (E).

The scale constructions were:

H: 9 point scale from 'very mechanical' to 'very human'

S: 10 point scale from 'very familiar' to 'very strange'

E: 10 points from 'not eerie' to 'extremely eerie'.

Controls

The task order was fixed for each participant so that the description task was always followed by the emotion task and then the final rating task. Within each task the order of presentation of the face types was randomised to control for order effects. While it was not possible to screen participants for naivety about the UVE, no explicit reference to the

theme of the study was made in the recruitment notices. None of the participants from Study 1 took part in Study 2.

Participants

212 participants were recruited to take part in the online study website hosted at the Open University. There were no special eligibility requirements so it was possible to issue an open invitation for participation which recruited members of staff at the institution where the study was carried out and social networks to which the researcher belonged. No direct contact was made to target individuals to ask them to take part and all participants volunteered to take part after reading messages with a link to the study which were posted on an institutional noticeboard, on an online blog or on social networking sites. While participants were not screened or selected for age, gender or location, they were asked if they would be willing to leave some basic demographic information but only at the end of the study once a full set of experimental data had been collected. 67% of the respondents to the study were female, most were aged between 21 and 40 and most reported living in the United Kingdom. A full breakdown of the responses to the age and location questions is shown in Appendix 1b.

Materials and Presentation

The stimulus materials in this online study were five black and white images of female faces. All five images were standardised to show the same portion of the face. They were also scaled to the same size of 320x240 pixels. This size was chosen after testing the web pages in several browsers and on several monitors and provided a clear image of the face while still allowing space in the rest of the page to present the rating scales and accompanying text.






| Artificial | Cluster 1 | Cluster 2 | Cluster 3 | Human |
|---|---|---|--|---|
| | Low human, highly strange and eerie | Medium human, strange and eerie | High human, low strange and eerie | |
|  |  |  |  |  |

Figure 9: Five face types presented to participants in this study.

The online study pages were designed to display the faces clearly and to present the relevant task in a way that participants could easily answer. Fonts and colours were chosen for ease of reading and high contrast. The pages used standard HTML code, did not require participants to have any plug-ins or non-standard components and were tested in a range of different web browsers to maximise the number of people who would be able to take part. The development of the study pages included a small pilot phase where colleagues tested the functionality of the pages in different browsers after which minor changes were made to the layout to improve clarity. The pilot testers did not take part in the final experiment.

The experimental part of the study was presented in four blocks, as detailed in the Procedure section. At all times during the study, participants were clearly told how many images they had seen so far and how many they still had to rate. This provided a count of images and also a percentage completion. Progress information was provided as a way to help participants remain motivated to work through all of the questions.

Participants' responses were stored in a Lotus Notes database hosted at the Open University. To ensure compliance with Data Protection legislation individual responses were only accessible to the researcher and to the database designer while the study was

live and once the final results had been passed to the researcher they were permanently deleted from the database. To ensure that responses to the survey were unique, IP address tracking was used so that if a participant came back to complete the survey from the same IP address they were told that their responses had already been collected.

Examples of the pages used in the study are given in Appendix 1c.

Procedure

Participants started the online study by clicking on a web link and landing on the study website. This consisted of three sections:

1) Introductory information and consent screen.

Once participants had arrived at the first page they were asked to read through an introduction giving details of what the tasks would involve. They were asked to give their consent to participate and were reassured of their right to withdraw at any point. They were then presented with the overarching scenario for the study. The three tasks were presented in the context of attending an imaginary conference. Participants were told:

'For the purposes of this study, I would like you to imagine that you are attending a robotics exhibition. You will be asked to carry out three different kinds of task relating to figures that you encounter while you are at the exhibition. I will then ask some questions about yourself, but these will only be used to analyse your answers by group characteristics, never to identify you personally.

Please give your answers honestly, and feel free to use as much detail as you wish. There are no right or wrong answers to the tasks.'

2) Task screens.

All task screens were laid out to present the image of the face in the upper right part of the screen, with the question on the upper left and the input area filling the lower part of the screen.

2A) Description task screens.

The scenario wording was: *'You arrive at the exhibition while the displays are still being set up. You start to look around while you are waiting for your colleagues to arrive. You see five figures that you would really like them to take a look at, but you know that they will be in a hurry when they get here. How would you describe this figure so that your colleagues will be able to find it easily?'*

A large free text area filled the lower part of the screen. There were no upper or lower limits on the numbers of characters that participants could enter.

2B) Emotion task screens.

The scenario wording was: *'During the exhibition, you learn that these figures are all prototypes for a new household robot. A researcher for the company who is building the robot would like to know how people would feel about having each of these in their home.'*

The free text area was presented as for 2a.

2C) Rating screens.

The question wording was: *'At the end of the exhibition, you are asked to complete an evaluation of the five figures you have been interested in. Please can you rate each of them on the following scales. Note that the scales may vary for each question. The scale points are not numbered, as we want you to think about feelings rather than precise numbers.'*

A radio-button scale was presented in the lower part of the screen. The scales were as described in the Design section: text labels were used at the ends of the scale but the individual points were not numbered. There were 15 rating screens in total presenting the five face types for the three scales.

3) Demographic Screens.

Once the experiment data collection was complete, participants were asked some questions about basic demographic information and invited to leave their contact details if they wanted to learn more about the project in the future.

Once participants had clicked a final button to save their answers they were taken to the main project website where they were given some further information about the UVT and the purpose of the study. This information was only available to participants who had completed all of the questions so was not available for the general public or potential participants to find by accident and potentially bias their responses to the survey.

Given that some of the faces used in the task had the potential to be unsettling, participants were invited to make contact with the researcher by email or telephone if they were disturbed or adversely affected at any point during their participation. While several emails were received from participants their content related to interest about the study or suggestions for future research directions rather than any areas of concern. It is therefore reasonable to conclude that no participants experienced adverse effects as a result of taking part and that from an ethical perspective the study did not give grounds for concern.

Content Analysis of free text data

The description and emotion tasks in this study were designed to collect rich and detailed qualitative comments about the different face types. In the main, the scenarios that featured in these tasks elicited lengthy responses from participants and some

participants even gave answers several paragraphs in length. In order to analyse these unstructured responses a content analysis of the unstructured data was performed. This coding approach was chosen over two other techniques, semantic resonance mapping and text characteristic analysis. Hand-coded content analysis was chosen as it offered the greatest level of engagement with the detail of the text data and offered the richest form of output. To do this, the dataset was manually hand-coded to allocate the comment itself to a designated category (e.g. 'Comment uses positive emotion terms') or to tag individual parts of the comment text to provide counts (e.g. 'Comment mentions eyes three times, mouth once.'). The categories and tags were designed to allow evaluation of the extent to which the faces were being processed analytically or holistically and whether the patterns of emotional response supported the idea of an UVE. For example, to explore whether participants approached the description task analytically or holistically, the number of times the following attributes were mentioned in comments were tagged and the number of instances counted: *Gender, Race, Age, Face as a whole, Eyes, Mouth, Nose, Hair, Skin, Chin, Cheek, Ear, Accessories*. References to these individual features could then be aggregated into whether the description was analytic or holistic in nature. To explore whether there were differences in participants reacting negatively, positively or neutrally across the different face types, comments were categorised according to those three dimensions, or put into a 'neither' or 'multiple' category where applicable. The categories were also designed to be as clearly defined as possible to allow reliable and accurate coding with several phases of checking built in to ensure that no comments had been mistakenly coded as the wrong category.

3.3.3: What will the subjective ratings of human-likeness, strangeness and eeriness be for an identified collection of NHAs?

3.3.3.1: Results

This section will present the results of analyses designed to explore this first research question by testing three hypotheses.

H1. Face type will have a significant effect on the ratings of human-likeness, strangeness and eeriness.

The ratings on each scale were analysed to compare five face types: the artificial and human anchor points were compared to the three exemplar faces, one drawn from each of the three clusters, as illustrated in Figure 9. Cluster 1 represented low human-likeness, high strangeness/eeriness. Cluster 2 represented medium human-likeness strangeness/eeriness and Cluster 3 represented high human-likeness, low strangeness/eeriness.

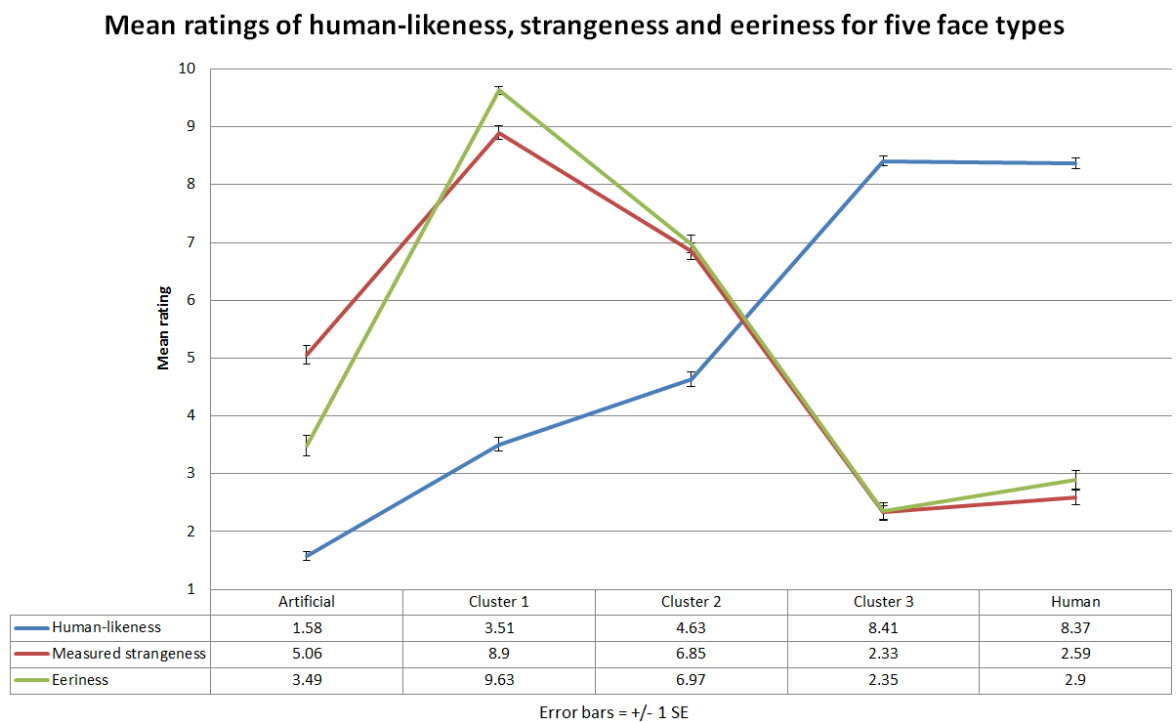


Figure 10: Comparison of the mean rating scales for each face type.

When the **human-likeness** ratings were analysed with a one-way within-participants ANOVA, Mauchley's test indicated that the assumptions of sphericity had been violated: $\chi^2(9)=49.47, p < .0005$, and so degrees of freedom were corrected¹ using the Huynh-Feldt estimates of sphericity ($\epsilon=.92$). The corrected results show that the ratings on the mechanical-human scale were significantly affected by the face type: $F(3.67,777.55)=1006.009, p < .0005$, partial $\eta^2 = .83$. Post-hoc paired t-tests using Bonferroni's correction (10 comparisons, p threshold=0.005) revealed significant differences between nine of the ten possible comparisons with only the Cluster 3-Human pair not showing a significant difference ($p=.74$).

When the **strangeness** ratings were analysed with a one-way within-participants ANOVA, Mauchley's test indicated that the assumptions of sphericity had been violated: $\chi^2(9)=32.71, p < .0005$, and so degrees of freedom were corrected using the Huynh-Feldt estimates of sphericity ($\epsilon=.94$). The corrected results show that the ratings on the eeriness scale were significantly affected by the face type: $F(3.75,792.16)=446.78, p < .0005$, partial $\eta^2 = .68$. Post-hoc paired t-tests using Bonferroni's correction (10 comparisons, p threshold=0.005) revealed significant differences between nine of the ten possible comparisons with the Cluster 3-Human pair not showing a significant difference ($p=.101$).

The **eeriness** ratings were analysed with a one-way within-participants ANOVA and Mauchley's test indicated that the assumptions of sphericity had been violated:

¹ In all instances where assumptions of sphericity are violated, the correction to the degrees of freedom was chosen based on the value of ϵ , following Field (2009). Where ϵ is over .75, the Huynh-Feldt correction has been used, otherwise the Greenhouse-Geisser correction has been used.

$\chi^2(9)=83.17, p < .0005$, and so degrees of freedom were corrected using the Huynh-Feldt estimates of sphericity ($\epsilon=.86$). The corrected results show that the ratings on the eeriness scale were significantly affected by the face type: $F(3.43,722.84)=472.53, p < .0005$, partial $\eta^2 = .69$. Post-hoc paired t-tests using Bonferroni's correction (10 comparisons, p threshold=0.005) revealed significant differences between eight of the ten possible comparisons with the Artificial-Human and Cluster 3-Human faces not showing a significant difference.

H2. When considering the ratings for each agent, the following patterns of correlation will be found:

- *Eeriness and strangeness will be strongly positively correlated*
- *Human-likeness and strangeness will be strongly negatively correlated*
- *Human-likeness and eeriness will be strongly negatively correlated*

The ratings were analysed using Spearman's r_s and support for the three hypotheses was found although the strength was not always as predicted. There was indeed a significant positive but moderately weak correlation between eeriness and strangeness ($r_s=.496, N=212, p<0.001$, one-tailed, explaining 25% of the variance.) A significant negative but weak correlation was found between human-likeness and strangeness ($r_s=-.262, N=212, p < 0.001$, one-tailed) and this only explained 7% of the variance. Finally, a significant negative correlation was found between human-likeness and eeriness ($r_s=-.224, N=212, p<.0001$, one-tailed). However, this is a weak correlation, explaining only 5% of the variance.

H3. When the mean human-likeness and strangeness ratings for each agent are plotted, the strangeness plot will show a deviation from a linear trend between the 50% and 100% human point.

This was not demonstrated in the results, as shown in Figure 11 below. The secondary curve approximates the relationship between strangeness and human-likeness that would be expected in an UVE where strangeness decreases with increasing human-likeness until the midpoint where strangeness sharply increases. It is an inverse version of Mori's curve which measured changes in familiarity as the scale used in Study 2's scale had strangeness at the highest end of that scale, rather than familiarity.

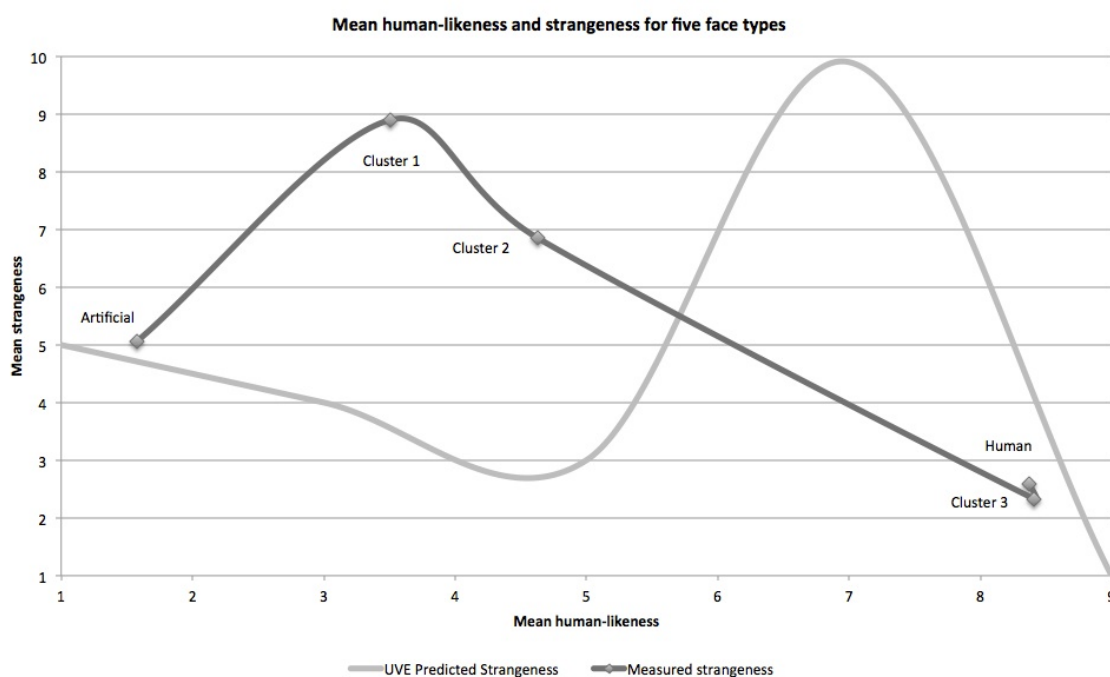


Figure 11: Plot of mean human-likeness against strangeness, overlaid with the curve that would be expected in an UVE.

3.3.3.2: Discussion

This initial exploration tested three hypotheses of whether the results of subjective ratings for human-likeness, strangeness and eeriness demonstrated a pattern indicative of the UVE. Firstly, tests were carried out into whether the different face types would be rated significantly differently in terms of their human-likeness, strangeness and eeriness. It was hypothesised that the face type (defined as either the cluster from which the exemplar face was drawn or the human/artificial anchor) would have a significant effect on each of the ratings, and also predicted patterns for the nature of this difference. (These

are the same correlations and plots predicted in Study 1 above and grounded in Mori's UV graph and MacDorman's (2006) scales derived from this, so the rationale for these predictions is the same.)

To explore H1, the findings of Study 1 predicted that the Cluster 1 face type would have low human-likeness, the Cluster 2 face type mid human-likeness and the Cluster 3 face type high human-likeness. As anchors deliberately chosen for their artificial or human nature, it would be expected that the Artificial face would be rated as the least human-like and the Human face type as the most human-like. The Human anchor face was rated overall as slightly *less* human-like than the Cluster 3 face with a mean rating of 8.37 rather than 8.41 but the post-hoc testing demonstrated that this difference was not statistically significant. It is possible that given the highly realistic nature of the Cluster 3 picture, participants may have had difficulty perceiving her as anything other than human.

The results of Study 1 predicted that the Cluster 1 face type would have high strangeness, the Cluster 2 face type mid strangeness and the Cluster 3 face type low strangeness. The anchor faces were not included in that study but as human faces should not appear strange to participants while robot faces may well be found to be strange, it could be expected that the human face would have a low strangeness rating and the robot face a moderate strangeness rating. As with the human-likeness ratings, the ratings found in this study did support this pattern with one small difference as Cluster 2 was rated as stranger than would have been predicted as a mean of 7 represents a high rather than mid rating. The Cluster 3 and Human face types were the only non-significant pair in the post-hoc analysis and the Human face was rated as slightly stranger than the Cluster 3 face but as was suggested in the human-likeness ratings, participants may have had

difficulty in making distinct ratings of strangeness for two faces that did appear similarly human-like.

The clusters derived from Study 1 were identified when human-likeness was plotted against strangeness rather than plotting it against eeriness but as eeriness was positively correlated with strangeness so it can be expected that the eeriness ratings should follow a similar pattern to that of strangeness, with high eeriness for Cluster 1, mid for Cluster 2 and low for Cluster 3. The results here do support that pattern.

Correlations between pairs of the measures were considered in this study to test H2, and while the predicted significant correlations were found in the expected directions, it was found that these correlations were generally weak. As with the results of the Study 1, the strongest of the three was a positive correlation between eeriness and strangeness, explaining 25% of the variance. The negative strangeness and human-likeness, and eeriness and human-likeness correlations explained only 7% and 5% of the variance respectively. It is therefore prudent to treat these results with caution, and even though they were found to be significant, to consider whether it would be valid to extrapolate beyond these findings to more general conclusions relating to the relationship between these variables.

The final hypothesis (H3) proposed that when the mean human-likeness and strangeness ratings for each face are plotted, the strangeness plot will show a deviation from a linear trend between the 50% and 100% human point. No such plot was found as the strangeness ratings peaked with the Cluster 1 faces, which were not rated as particularly close to human. The Human and Cluster 3 faces were both rated low for eeriness. However, the Cluster 2 face had been identified as a likely candidate for an uncanny response based on a review of MacDorman's (2006) results and the results from

this study did support those criteria: the Cluster 2 face was rated having a mean eeriness of over 4.5 for human-likeness as well as a mean eeriness and strangeness of over 6.

3.3.4: When asked to describe the faces of five NHAs, which features will be emphasised and how will this compare to the descriptions of human and non-human comparison faces?

3.3.4.1: Results

This section will explore two hypotheses about how participants would describe the different face types.

H4. When descriptions are analysed for the frequency of types of descriptive term, there will be a significant difference between the frequency of terms describing the whole face compared to those describing individual features. The rationale for this hypothesis is to begin to explore whether there is a relationship between levels of human-likeness and how faces are processed and particularly whether there was any evidence that NHFs were processed analytically rather than holistically. At this stage, while this was an area of interest, there was no suggestion from the supporting literature that there would be a specific pattern of response by human-likeness, so no specific predictions were made about the nature of that difference.

The free text descriptions about each face type were analysed to count how frequently key terms were used. These were either specific facial features (e.g. eyes, nose, mouth, chin) or general attributes about the agent (e.g. female, middle-aged, Caucasian). The frequency of terms used to describe each face type were aggregated to give totals of analytic and holistic descriptions. The sum of references to eyes, nose, mouth and chin formed the analytic description total and the sum of references to age, gender, ethnicity,

or the word ‘face’ created the holistic description total. Table 5 presents the means and standard deviations of the frequencies of each key term for the five face types.

| Face type | | Artificial | | Cluster 1 | | Cluster 2 | | Cluster 3 | | Human | |
|----------------|-----|------------|----------------|-----------|----------------|-----------|----------------|-----------|----------------|-------|----------------|
| Term | N | Mean | Std. Deviation | Mean | Std. Deviation | Mean | Std. Deviation | Mean | Std. Deviation | Mean | Std. Deviation |
| Eyes | 212 | 1.18 | 0.83 | 1.09 | 0.60 | 0.87 | 0.73 | 0.28 | 0.53 | 0.65 | 0.75 |
| Mouth | 212 | 0.38 | 0.50 | 0.40 | 0.55 | 0.50 | 0.69 | 0.42 | 0.59 | 0.64 | 0.66 |
| Nose | 212 | 0.05 | 0.21 | 0.34 | 0.49 | 0.52 | 0.52 | 0.05 | 0.22 | 0.20 | 0.46 |
| Chin | 212 | 0.96 | 0.19 | 0.83 | 0.41 | 0.67 | 0.58 | 0.83 | 0.41 | 0.33 | 0.58 |
| Analytic total | 212 | 1.84 | 1.23 | 1.85 | 1.33 | 1.90 | 1.43 | 0.77 | 1.05 | 1.50 | 1.41 |
| Face | 212 | 0.61 | 0.78 | 0.31 | 0.56 | 0.60 | 0.72 | 0.16 | 0.39 | 0.53 | 0.63 |
| Gender | 212 | 0.07 | 0.26 | 0.25 | 0.43 | 0.51 | 0.50 | 0.76 | 0.44 | 0.72 | 0.47 |
| Race | 212 | 0.00 | 0.00 | 0.02 | 0.14 | 0.08 | 0.26 | 0.72 | 0.46 | 0.19 | 0.39 |
| Age | 212 | 0.08 | 0.33 | 0.20 | 0.41 | 0.05 | 0.23 | 0.21 | 0.42 | 0.27 | 0.51 |
| Holistic total | 212 | 0.75 | 0.85 | 0.77 | 0.86 | 1.23 | 0.91 | 1.85 | 0.94 | 1.71 | 1.08 |

Table 5: Mean frequencies of references to key terms for five face types.

These totals were analysed using a two-way within-participants ANOVA. A 5 x 2 design was employed with five levels of the face type factor (Artificial, Cluster 1, Cluster 2, Cluster 3 and Human) and two levels of description type factor (analytic and holistic). The interaction between face type and description type was significant: $F(3.8, 801.5) = 89.78$, $p < .0005$, partial $\eta^2 = .29$. Sphericity had been violated for this interaction: $\chi^2(9) = 33.1$, $p < .0001$, $\epsilon = .95$ so Huynh-Feldt corrections were used for the degrees of freedom. For the main effect of face type, Mauchley's test indicated that the assumption of sphericity had been violated: $\chi^2(9) = 26.1$, $p < .0001$, $\epsilon = .96$. Using corrected degrees of freedom according to the Huynh-Feldt estimates of sphericity, the results showed a significant effect of face type: $F(3.83, 809.7) = 12.98$, $p < .0005$, partial $\eta^2 = .06$. The main effect of description type was also significant: $F(1, 211) = 20.74$, $p < .0005$, partial $\eta^2 = .09$.

The interaction graph in Figure 12 has been produced to illustrate the relationship between the variables.

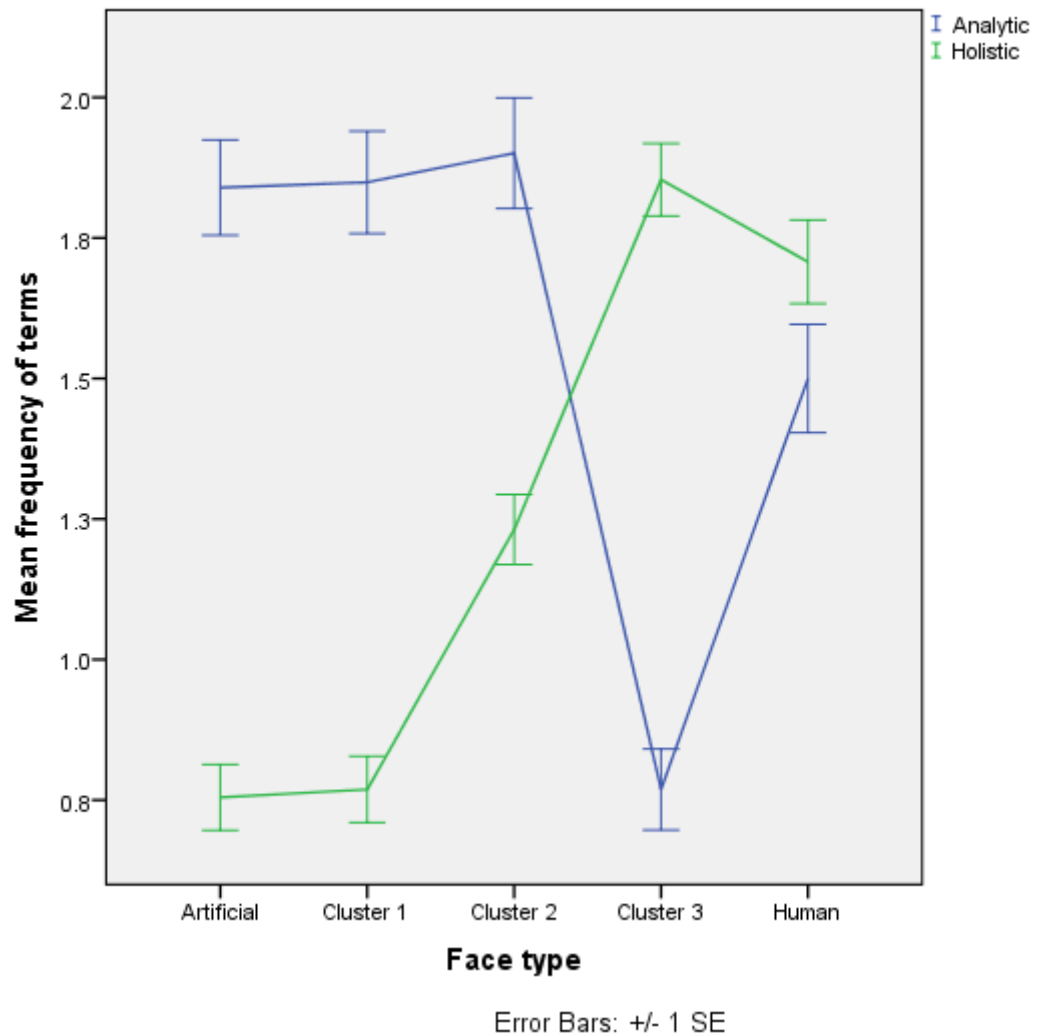


Figure 12: Holistic terms were most frequently used when describing the Human and Cluster 3 faces, analytic terms for the Cluster 2 face.

Artificial and Cluster 1 faces were described using mainly analytic terms with the other three face types showing different patterns where a mix of analytic and holistic terms were used. Participants used a lot of analytic terms to describe the Cluster 2 face, but holistic terms were used as well. The Cluster 3 face showed the inverse pattern to the Artificial and Cluster 1 faces, with considerably more holistic descriptions than analytic descriptions. The human face had a high number of holistic terms, as was expected, but also a high number of analytic terms.

H5. When the descriptions are analysed for the frequency of references to different facial features, the eyes will be mentioned most frequently for the moderately human face (Cluster 2).

There was a significant difference between the face types: when analysed with a one-way within-participants ANOVA, Mauchley's test indicated that the assumptions of sphericity had been violated: $\chi^2(9)=47.13$, $p < .0005$, and so degrees of freedom were corrected using the Huynh-Feldt estimates of sphericity ($\epsilon=.941$). The corrected results show that face type was a significant factor in the frequency of descriptions of eyes: $F(3.766,722.84)=76.58$, $p < .0005$, partial $\eta^2=.266$. Figure 13 below shows the mean number of references to eyes for each face type.

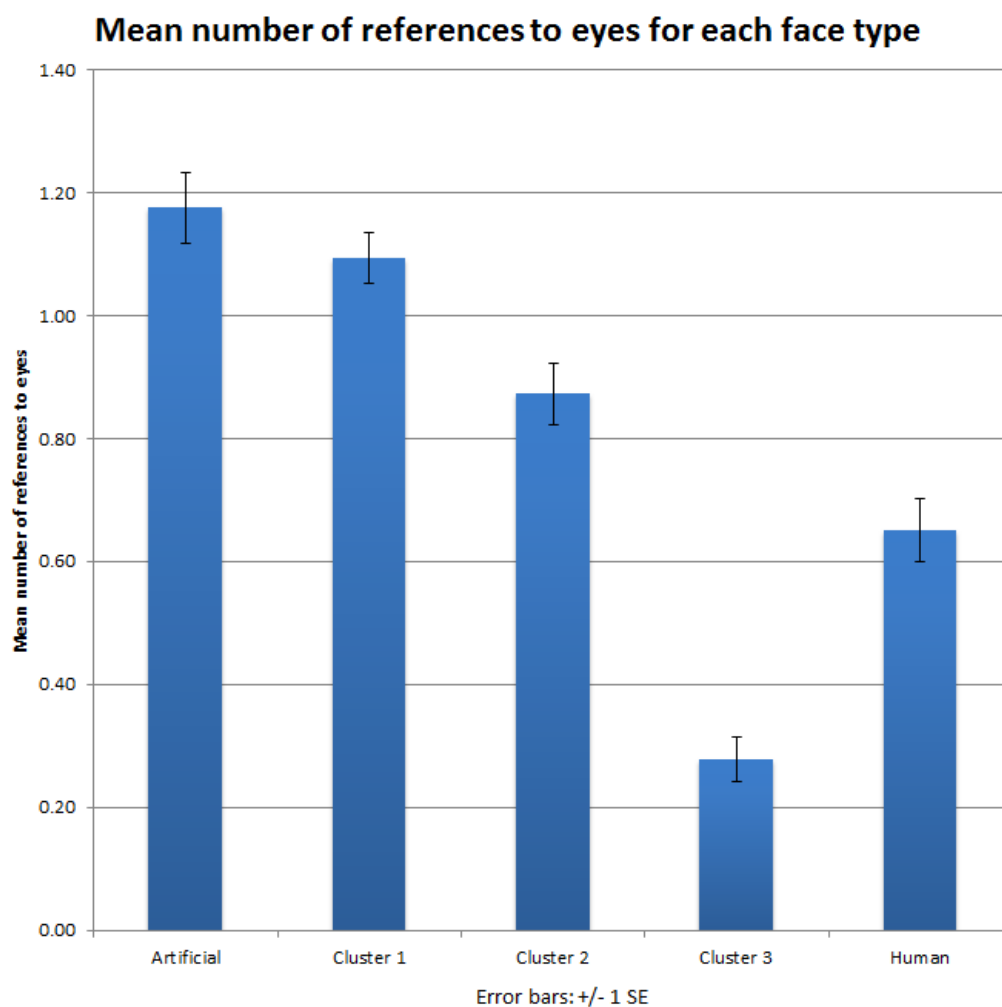


Figure 13: Mean number of references to eyes used when describing the five different face types.

Examination of the repeated contrasts for face type found significant differences between three of the four possible comparisons with only the Artificial and Cluster 1 pair emerging as non-significant. ($p=.212$) (Cluster 1 to Cluster 2: $F(1,211)=13.20$, $p < .0005$, partial $\eta^2=.059$, Cluster 2 to Cluster 3: $F(1,211)=130.47$, $p < .0005$, partial $\eta^2=.38$, Cluster 3 to Human: $F(1,211)=54.70$, $p < .0005$, partial $\eta^2=.21$)

It can be concluded that there was a significant difference in the number of references to eyes between all the face types except for the artificial and the Cluster 1 face, which was the low-human, high-eeriness face. There was no significant difference in the number of eye references between those clusters. However, the hypothesis that eyes would be mentioned most often for the moderately human face drawn from Cluster 2 was not supported: the face type with the highest frequency of references to eyes was Artificial ($M=1.18$) rather than the hypothesised moderately-human Cluster 2 face ($M=0.87$).

3:3:4:2: Discussion

The analyses in this section explored whether there was any evidence from the open comments to demonstrate whether participants used generally holistic or analytic terms to describe the different faces, based on open comments given by participants in response to a question asking how they would approach the task of describing each agent.

H4 explored which features would feature prominently in the descriptions of the NHFs and how they would compare with the features used to describe the human and non-human faces. H4 quantified this question to explore whether there would be a significant difference between the frequency of terms describing the whole face compared to those describing individual features. Evidence that participants were focused particularly on

certain features, or were generally using descriptions of facial features rather than of an impression of the overall face when describing the NHFs compared to the human or non-human faces would support the theory that human-likeness has an effect on how faces are processed. No specific predictions were made about the nature of that difference as although anecdotal evidence suggests certain patterns of concentration, there was no suggestion from the supporting literature relating the UVE to patterns of human-likeness, eeriness and analytic processing, so no specific predictions were made about the nature of that difference.

A significant interaction was found between face type and the type of description term used. Analytic terms were used most for the Artificial, Cluster 1 and Cluster 2 faces, and were used least for the Cluster 3 faces with a moderate number used for the Human faces. Holistic terms were used most for the Cluster 3 and Human faces, least for the Artificial and Cluster 1 faces, and moderately used for Cluster 2. It is possible to separate the face types into two groups in terms of term use as the Artificial, Cluster 1 and Cluster 3 faces were described in either analytic or holistic terms but the Human and Cluster 2 terms did not show such a clear distinction between the terms used. Of the three images, the Cluster 2 face was the one identified as the potential UV candidate image as it sat in a region of medium strangeness and eeriness and medium human-likeness, so this finding could be due to UV images being harder to describe and that there was less of a consensus in the approach taken to describe the faces. However, the human face was also described in both analytic and holistic terms, and as this was the face introduced as an anchor for comparison this was an unexpected finding. It should be noted that these findings may have been due to the specific characteristics of the images that were used in the study.

H5 explored whether the UV candidate face drawn from Cluster 2 faces would have the highest number of references to eyes when compared to the other clusters. This hypothesis is grounded in the idea that it may be an aspect of how eyes are presented that can make NHFs eerie. The Cluster 2 image, as a moderately human and moderately strange/eerie face was drawn from the cluster containing the most likely UV candidates, and so it was predicted that the references to eyes would be highest for this cluster. While a significant difference in the number of references to eyes was found in the results, that particular hypothesis was not supported as eyes were mentioned most frequently for the Artificial face.

The method used in this study was to review the comments made by participants and allocate them to different categories, which could then be used for analysis to see whether there were significant differences in term usage for each of the five faces. While this is a valid and useful way of distilling a large body of qualitative data to allow comparisons to be drawn, it is useful to be critically reflective at this stage about how the comments were allocated. In designing this study, choices were made in how references should be allocated to individual categories and then in how those should be grouped further into the over-arching categories of whether each face was being perceived in a broadly analytic or holistic manner. It could be argued that some of the categories selected as evidence of holistic processing (for example, race) could equally have been allocated to the analytic category as they can be used as a way to convey a perception of aspects of the face. Different design choices may have led to some different conclusions at the overarching category level. This study has also taken the view that the approaches participants use to describe or react to the images they see can be used through this process of distillation to extrapolate indicators of how the images are actually being processed. It is acknowledged that this involves several steps that must have occurred

between the way that the image was actually processed and the participant's ability to reflect this in their behaviour through their comments. However, the studies that follow in Phase 2 and Phase 3 will take a more rigorous and experimental approach to build on these findings and ensure that more direct, and less exploratory, conclusions can be drawn.

These findings may be indicative of differences between how people process human, near-human, and artificial faces but as this study examined one example of each face type so further research would be required to ascertain whether such processing differences can be generally applied to NHAs.

3.3.5: When considering a hypothetical social encounter with the NHAs, what terms will be used to express how people would feel about this and which emotions will characterise their reactions?

3.3.5.1: Results

This section will explore two hypotheses about social encounters with NHAs.

H6. The response descriptions will be analysed for how frequently positive, negative and neutral emotion terms are used to describe each face:

- a. The highest frequency of negative emotions will be used for the moderately human face type.*
- b. The lowest frequency of positive emotions will be used for the moderately human face type.*
- c. There will be no significant difference in the number of neutral emotion terms used for each face type.*

Participant's comments about how they would feel sharing their home with each of the NHAs were categorised according to key terms relating to emotions. These were first

coded as references to individual emotions (e.g. happy, angry, excited, sad) and then aggregated into categories of positive, negative or neutral. Table 23, Table 24, and Table 25 in Appendix 1d present the frequencies of individual positive, negative and neutral terms used for each face type and Table 6 below shows the mean and standard deviation for the aggregate categories of each emotion.

| Face type | | Artificial | | Cluster 1 | | Cluster 2 | | Cluster 3 | | Human | |
|-----------|-----|------------|----------------|-----------|----------------|-----------|----------------|-----------|----------------|-------|----------------|
| Emotion | N | Mean | Std. Deviation | Mean | Std. Deviation | Mean | Std. Deviation | Mean | Std. Deviation | Mean | Std. Deviation |
| Positive | 212 | 0.37 | 0.52 | 0.01 | 0.12 | 0.08 | 0.27 | 0.58 | 0.79 | 0.34 | 0.55 |
| Negative | 212 | 0.21 | 0.49 | 1.00 | 0.72 | 0.65 | 0.66 | 0.15 | 0.37 | 0.27 | 0.49 |
| Neutral | 212 | 0.08 | 0.27 | 0.00 | 0.07 | 0.08 | 0.28 | 0.07 | 0.26 | 0.12 | 0.32 |

Table 6: Mean frequencies of references to key emotion terms for five face types.

Figure 14 depicts the means of the positive, negative and neutral frequencies for each face type as a bar chart for ease of comparison.

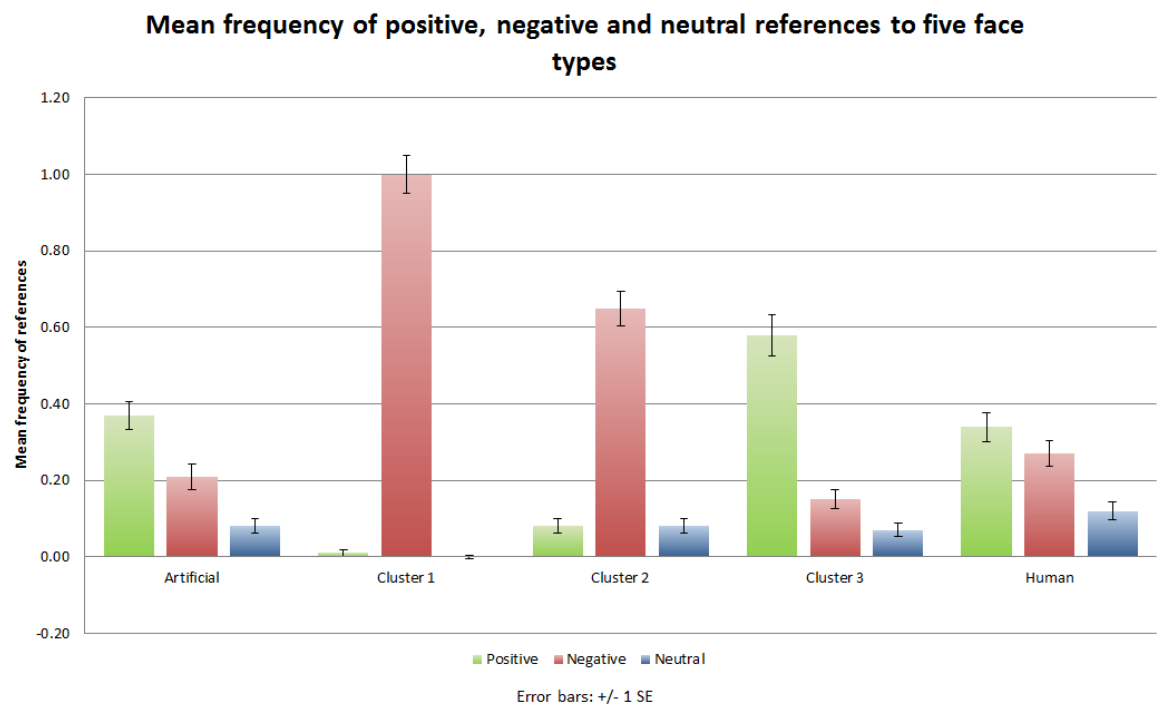


Figure 14: Bar chart comparing mean frequencies of references to key terms for five face types.

To investigate H6a, the frequency of negative emotions was analysed using a one-way within-participants ANOVA. There was a significant difference between the face types. When analysed with a one-way within-participants ANOVA, Mauchley's test indicated that

the assumptions of sphericity had been violated: $\chi^2(9)=59.70$, $p < .0005$, and so degrees of freedom were corrected using the Huynh-Feldt estimates of sphericity ($\epsilon=.90$). The corrected results show that face type was a significant factor in the frequency of negative emotions used to describe each face type: $F(3.61,762.03)=100.93$, $p < .0005$, partial $\eta^2=.32$. Examination of the repeated contrasts for face type found significant differences between all of the four possible comparisons: Artificial to Cluster 1: $F(1,211)=217.79$, $p < .0005$, partial $\eta^2=.51$, Cluster 1 to Cluster 2: $F(1,211)=38.75$, $p < .0005$, partial $\eta^2=.16$, Cluster 2 to Cluster 3: $F(1,211)=85.37$, $p < .0005$, partial $\eta^2=.29$, Cluster 3 to Human: $F(1,211)=10.74$, $p < .0005$, partial $\eta^2=.05$. However, the face type with the highest frequency of negative references was Cluster 1 ($M=1.0$) rather than the hypothesised moderately-human Cluster 2 face ($M=0.65$).

To investigate H6b, the frequency of positive emotions was analysed using a one-way within-participants ANOVA. There was a significant difference between the face types. Mauchley's test indicated that the assumptions of sphericity had been violated: $\chi^2(9)=310.6$, $p < .0005$, and so degrees of freedom were corrected using the Greenhouse-Geisser estimates of sphericity ($\epsilon=.66$). The corrected results show that face type was a significant factor in the frequency of positive emotions used to describe each face type: $F(2.66,562.1)=48.24$, $p < .0005$, partial $\eta^2=.18$. Examination of the repeated contrasts for face type found significant differences between all of the four possible comparisons: Artificial to Cluster 1: $F(1,211)=94.62$, $p < .0005$, partial $\eta^2=.31$, Cluster 1 to Cluster 2: $F(1,211)=12.9$, $p < .0005$, partial $\eta^2=.06$, Cluster 2 to Cluster 3: $F(1,211)=77.12$, $p < .0005$, partial $\eta^2=.27$, Cluster 3 to Human: $F(1,211)=18.14$, $p < .0005$, partial $\eta^2=.08$. However, the face type with the lowest frequency of positive references was Cluster 1 ($M=0.1$) rather than the hypothesised moderately-human Cluster 2 face ($M=0.08$).

To investigate H6c, the frequency of neutral emotions was analysed using a one-way within-participants ANOVA. There was a significant difference between the face types which did not support the original hypothesis which predicted that while the number of positive and negative emotions would vary, there would be no difference for the neutral emotions: Mauchley's test indicated that the assumptions of sphericity had been violated: $\chi^2(9)=72.5, p < .0005$, and so degrees of freedom were corrected using the Greenhouse-Geisser estimates of sphericity ($\epsilon=.87$). The corrected results show that face type was a significant factor in the frequency of neutral emotions used to describe each face type: $F(3.49,737.65)=6.31, p < .0005$, partial $\eta^2=.29$. Examination of the repeated contrasts for face type found significant differences between two of the four possible comparisons: Artificial to Cluster 1 ($F(1,211)=15.13, p < .0005$) and Cluster 1 to Cluster 2 ($F(1,211)=16.13, p < .0005$). (Cluster 2 to Cluster 3 and Cluster 3 to Human were not significant, $p=.55$ and $p=.04$ respectively.)

H7. Emotion terms synonymous with eeriness or unease will be highest for the moderately human face type.

To explore this hypothesis, the keywords used in each comment were examined to derive a count of the number of references to eeriness, unease or a synonymous term for each face type². Table 7 below shows the means and standard deviations for the frequency of eeriness synonyms for each face type.

² These were: creeped/creeps/creepy; disconcerted; disturbed/disturbing; eerie; freaked/freaky; frightened/frightening; haunted; nervous; nightmarish; scared; spooked; uneasy; unnerved; unsettled.

| Face type | N | Artificial | | Cluster 1 | | Cluster 2 | | Cluster 3 | | Human | |
|-------------------|-----|------------|----------------|-----------|----------------|-----------|----------------|-----------|----------------|-------|----------------|
| | | Mean | Std. Deviation | Mean | Std. Deviation | Mean | Std. Deviation | Mean | Std. Deviation | Mean | Std. Deviation |
| Eeriness synonyms | 212 | 0.08 | 0.28 | 0.61 | 0.49 | 0.30 | 0.46 | 0.08 | 0.28 | 0.11 | 0.32 |

Table 7: Mean references to terms synonymous with eeriness for each face type.

When analysed using a one-way within-participants ANOVA, there was a significant difference between the face types in terms of the number of references to eeriness or synonymous terms. Mauchley's test indicated that the assumptions of sphericity had been violated: $\chi^2(9)=105.78$, $p < .0005$, and so degrees of freedom were corrected using the Greenhouse-Geisser estimates of sphericity ($\epsilon=.80$). The corrected results show that face type was a significant factor in the frequency of eeriness synonyms used to describe each face type: $F(3.23,681.82)=81.68$, $p < .0005$, partial $\eta^2=.28$. Examination of the repeated contrasts for face type found significant differences between three of the four possible comparisons: Artificial to Cluster 1: $F(1,211)=195.01$, $p < .0005$, partial $\eta^2=.48$, Cluster 1 to Cluster 2: $F(1,211)=46.17$, $p < .0005$, partial $\eta^2=.18$, Cluster 2 to Cluster 3: $F(1,211)=38.99$, $p < .0005$, partial $\eta^2=.16$. The comparison of Cluster 3 to Human was not significant, $p=.26$.) However, the face type with the most references to terms synonymous with eeriness was the Cluster 1 face, rather than Cluster 2 as hypothesised.

3.3.5.2: Discussion

This section of the study used measures derived from the open comments given by participants about their feelings on potential social encounters with NHAs. As such it represents an exploration of emotional responses, and serves as an indication for further research. It used two hypotheses to explore the terms that participants would use to describe how they might feel in an encounter with an NHA, and also explored specified patterns of response terms.

The first of the two hypotheses explored in this section was that three patterns would be found in terms of the type of emotion terms used for each face type. The rationale for these predictions is that the moderately human face was drawn from the cluster containing the three best UV candidates and the concept of the UV requires that there is a negative or unsettling emotional component to the response to NHFs which would predict that the moderately human face would be described more negatively and less positively.

H6 was explored in three parts:

a. The highest frequency of negative emotions will be used for the moderately human face type.

A significant difference was found here, but the highest number of negative descriptions was actually found for the low human-like/highly strange and eerie (Cluster 1) face rather than the moderately human face.

b. The lowest frequency of positive emotions will be used for the moderately human face type.

The finding above was mirrored for the lowest frequency of positive emotions, with the Cluster 1 face receiving the lowest frequency of positive emotions. While it may be a common-sense assumption that these two would represent opposites in the analysis, it is still a reassuring finding that the results do indeed bear this out, and it adds validity to the measurement scales and tasks that were employed in the study.

c. There will be no significant difference in the number of neutral emotion terms used for each face type.

This hypothesis predicted that neutral terms would be used equally for all of the face types, but it was found that this was not the case. More neutral terms were used for the human face than any other.

H7 was a direct exploration of the linguistic terms that were used by participants in describing the different faces and served as a method for testing the extent to which the different face types were being described in terms synonymous with eeriness and unease. It serves as a more direct check of whether the UV candidate face type was actually described as more eerie when viewed by participants. As the moderately human cluster (Cluster 2) was predicted to contain the most likely candidate images for eliciting an UVE, it was predicted that terms synonymous with eeriness would be most frequent for that face. Emotion terms synonymous with eeriness or unease will be highest for the moderately human face type. Following the findings above, a significant difference was found but it was the Cluster 1 face where participants used terms synonymous with eeriness to describe the face.

3.3.6: How will the ratings of eeriness relate to the identified features and emotions?

3.3.6.1: Results

The results so far have considered aspects of the scale-based ratings of human-likeness, strangeness and eeriness as well as the methods used by participants in describing each face type and saying how they would feel about sharing their home with the different NHAs. The final results section for this chapter will consider the eeriness measure in more detail: as it is core to the UVE theory, three potential relationships between eeriness and other measures derived in this study will be considered using correlational methods.

H8. There will be a correlation between ratings of eeriness and the frequency of descriptions referring to individual features rather than to the whole face.

H9. There will be a positive correlation between ratings of eeriness and the frequency of references to eyes.

H10. Eeriness will be positively correlated with the proportion of descriptions containing a word synonymous with uncanniness.

These three hypotheses were explored using Spearman's r_s correlations. The results of the correlations are summarised in Table 8 below.

| Face type | N | Number of references to individual features ⁺ | | Number of references to eyes | | Proportion of terms synonymous with uncanniness | |
|------------|-----|--|------|------------------------------|------|---|------|
| | | r_s | p | r_s | p | r_s | p |
| Artificial | 212 | .000 | .995 | -.008 | .456 | 0.388** | .000 |
| Cluster 1 | 212 | -.015 | .825 | -.026 | .352 | 0.211** | .001 |
| Cluster 2 | 212 | -.121 | .078 | -0.162** | .009 | 0.263** | .000 |
| Cluster 3 | 212 | .039 | .567 | .065 | .172 | 0.300** | .000 |
| Human | 212 | .016 | .817 | -.070 | .156 | 0.315** | .000 |

+ = 2-tailed test

* = Correlation is significant at the 0.05 level (1-tailed).

** = Correlation is significant at the 0.01 level (1-tailed).

Table 8: Summary of correlation results for H8, H9 and H10

Six of the fifteen correlations were found to be significant. It can be seen that only H10 was fully supported across each of the five face types, with the proportion of terms synonymous with uncanniness being significantly and positively correlated with eeriness. However, these correlations were not strong, with the highest for the Artificial cluster explaining only 15.05% of the variance. The other significant correlation related to H9, as a significant negative correlation was found between the number of references to eyes and eeriness for the Cluster 2 face type.

3.3.6.1: Discussion

The final exploration in this study looked across the different tasks to understand the relationship between eeriness ratings for each face type and the measured frequency of references to individual features as opposed to whole face descriptors, references to eyes, and emotional responses containing words synonymous with eeriness.

Firstly, H8 predicted that a significant correlation would be found between ratings of eeriness and the frequency of descriptions referring to individual features rather than to the whole face. The rationale for this question is to test the concept that eerier faces are processed analytically rather than holistically, as described in H4, but the difference with this hypothesis is that it uses the eeriness ratings for each face rather than making a prediction based on the moderately human face type as the potential UV candidate. This was explored as a two-tailed hypothesis as while a difference was expected, the literature did not support a prediction of the nature of this difference. However, this hypothesis was not supported as no significant relationships were found between eeriness and term usage for any of the face types.

The concept was explored in more detail in H9 which predicted that there would be a positive correlation between ratings of eeriness and the frequency of references to eyes: this develops the concept detailed in H5 above that one of the explanations for the eeriness in the UVE may be some aspect of eye appearance where higher ratings of eeriness are linked to referring to the eyes more in the description. These were only significantly correlated for one face type, Cluster 2. However, this was a weak negative correlation rather than a positive one as predicted.

Finally, in H10 it was hypothesised that ratings of eeriness would be positively correlated with the proportion of descriptions containing a word synonymous with uncanniness. The rationale here was to examine whether the eeriness ratings were in line with the emotional response behaviour as it would be expected that as rated eeriness increases, so would the emotional response behaviour. Support was found for this hypothesis, with a significant and positive correlation found for each of the five face types. This helps to provide evidence that the eeriness scale is robustly measuring that

aspect of the UVE and adds to the reliability of this scale as an instrument for investigating uncanniness.

The implications for these findings in terms of the UV and face perception literature will now be considered in a general discussion of the studies which have been carried out in this first phase of research.

3.4: General Discussion of Phase 1

Phase 1 started with the basic concept of Mori's UVE and then looked at how it could be operationalised to see whether there was indeed a relationship between the human-likeness of an agent and how people viewing that agent would respond. It was designed to address a knowledge gap in that although a UVE could be elicited as a function of increasing human-likeness, there was insufficient psychological evidence explaining why it occurred or a detailed description of the emotional response itself. This study collected quantitative and qualitative data as a first step in answering this question, posed by Pollick (2009):

'what bit of the human response to these artefacts is the essential aspect of its uncanny nature?' (p76)

Similarly, Steckenfinger and Ghazanfar (2009) asked which aspects of the appearance of NHAs elicit our expectations of human appearance and behaviour but then fail to live up to those expectations. As described in the design principles for investigating the UVE, (Table 1 in Section 3.1) the first step towards doing this is to operationalise the key variables of human-likeness, strangeness and eeriness and test these with examples of NHAs, which could potentially display uncanny properties. It was possible to predict the nature of the relationship between those variables from the UVE literature described in Section 1.1, 1.2 and 1.3 and to test whether NHAs would elicit corresponding ratings of

strangeness and eeriness. Ratings of human-likeness, strangeness and eeriness were collected for twenty NHFs, from which three images were identified for further exploration, along with two similar anchor images. Correlations and visualisations were used to examine the rating results and limited support for the UVE was found as the correlations indicated that the expected patterns of relationship did occur. Eeriness and strangeness were always positive correlated, and both were negatively correlated with human-likeness. However, when the mean results were plotted against each other to emulate Mori's graph, the classic 'valley' in familiarity did not appear in either case.

In Study 1, the NHFs formed three distinct clusters of images rated as highly, moderately or barely human-like, with corresponding ratings of barely, moderately and highly strange. These three clusters mirror Tinwell's (2009) finding in a similar study where a range of different NHAs were rated on human-likeness and familiarity and rather than a single 'uncanny' dip, three distinct zones of high strangeness were found as human-likeness increased. One possible explanation for Tinwell et al's multiple valleys and the image clusters found in Study 1 here may have been the lack of human and non-human reference faces in the sets of images to be rated, as having no point of comparison for the NHAs meant that it was not possible to be sure that the human-likeness continuum being mapped spanned all the full range of possible points. Study 2 in this thesis added human and artificial faces along with the NHFs to address this issue by testing an anchored range of responses by using human, three types of near-human and non-human faces, but the typical UVE was not found. Cluster 2 was predicted to be the face that would fall into a 'valley', where it was rated as moderately human and strange/eerie, but this was not found. The premise for that prediction was that it represented Study 1's cluster of potentially unsettling faces so either it was not the best exemplar of that type, or it may be that those qualities did not generalise beyond the scope of Study 1. This relates back to

this issue of circularity discussed in Section 2.3.1 which presents a methodological difficulty in any exploration of the UVE. The three NHFs selected for Study 2 were chosen subjectively from the faces in each cluster, so the inclusion of anchor faces in a stimulus image set alone may not be sufficient to avoid this circularity issue.

The collection and analysis of the subjective rating data in both studies served to ground the stimulus images in terms of their human-likeness and strangeness, but did not provide any information about the nature of participants' responses to them. The second study explored this question by asking participants to describe the five faces and also to imagine how it would be to spend time in close proximity with an agent with that face. The aim was to begin to gather unstructured data on the broad question of how people describe different faces and also how their emotional reactions to each face would vary. The measurements used for this study were based on a quantification of the free text responses to explore specific hypothesis drawn from the UVE literature and anecdotal observation of faces which were subjectively judged to be eerie. One point to note is that several of the comments that have been coded as holistic were ones where participants compared the stimulus image to famous people, characters from video games or compared them to racial archetypes.

Participants used more terms relating to individual features to describe the eerier faces, compared to the human and highly human face in Cluster 3, and they also used terms relating to eyes more. The table below presents several of the comments that referenced the eyes of the stimulus images, and the cluster to which they belonged.

| Comment | Image |
|--|------------|
| A[n] android that seems to have been designed to be a cartoon but its cold huge eyes are too focused and direct. This causes the viewer to become very self conscious and paranoid. The pale "skin" is also a problem because it makes the android seem undead | Artificial |
| a wee "toy story" style head who looks a wee bit sad and has surprised eyebrows. No ears. And has a webcam in his forehead, framed with squares. | Artificial |
| Cartoonish 'Toy Story' type head, androgynous figure, with large eyes, jointed jaw and hole in middle of forehead surrounded by small squares forming oblong shape. | Artificial |
| Obviously artificial, large eyes that appear to be movable, single-piece separate jaw, rectangle in forehead with circle in middle, looks like a refugee from a Daft Punk video. | Artificial |
| A terrifying girl-child doll with no eyes. | Cluster 1 |
| Disturbing small child with huge empty eye sockets, tiny pert mouth, heart shaped face, tiny nose, and hair with straight bangs. | Cluster 1 |
| Freaky Japanese doll face with large black pits instead of eyes. Straight hair with heavy fringe. | Cluster 1 |
| Ghostly, slightly unsettling and creepy. A dollish girl's face with blank hollows for eyes, a small, upturned nose and a pinched little mouth. | Cluster 1 |
| The one that looks like she's got clay or something for a nose, and dead eyes | Cluster 2 |
| A face that looks like a living statue with eyes that look like pools of water. | Cluster 2 |
| CGI face with dodgy eyes and pursed lips | Cluster 2 |
| It's a really disturbing female face that looks a lot like a normal face got smeared over somehow. It looks worried and it's got the weirdest nose and these gel-like eyes. | Cluster 2 |
| It's the one that looks slightly melted, with the eyes unevenly shaped and a nose that looks kind of like dripping candle wax. | Cluster 2 |
| This face looks almost sad and of a futuristic quality. They have big eyes and their mouth is in a neutral expression. | Cluster 2 |
| This robot, it has large moist lips, a thin nose and large light eyes. Whilst it's features are humanoid, it looks slightly unreal, like a composite of FHM's ideal woman. | Cluster 2 |
| Lifelike human female with butterfly earrings. Something slightly off about the hair and eyes. | Cluster 3 |
| This is a human face with no obvious, outstanding features. The left eye is smaller / more closed than the right. The figure has a Mona Lisa smile. | Human |

Table 9: Verbatim comments where references were made to eyes in a creepy or unsettling context

To an extent this had been predicted from the anecdotal evidence that 'uncanny' faces were often characterised by unusual or distorted eyes but since the design of this phase, more recent research (Seyama and Nagayama (2007, 2009) and Looser and Wheatley (2010)) has also found support for the role of eyes in the UVE. Study 2 was an exploratory study with the aim of directing later research so this finding was not taken as a firm conclusion confirming the role of eyes in eliciting the UVE but rather as corroboration for the hypothesis that this may be part of the effect and an indication of an area for the next phase of research.

The last part of this phase looked at the emotions that were elicited when participants described how they would feel sharing their home with an android possessing the

appearance of each of the NHAs. This task was designed to see whether there was any evidence that eeriness was elicited for the NHFs in Clusters 1-3, and to see whether there was any systematic pattern to the types of emotions that were used to describe each face. The measurements were again drawn from analyses of unstructured comments given by participants, and found support for the hypotheses that negative terms were used most for the eeriest face.

It is possible to draw tentative conclusions from the research questions posed in this first phase, and these findings have suggested areas for follow-up studies. In reviewing Study 1 and 2, some limitations were noted. A key point to note from these studies is the importance of ensuring that reference images are always included when working with NHFs to serve as controls for rating data and points of comparison when drawing conclusions about the nature of the UVE. Without a gradual progression in human-likeness it is difficult to confidently locate any observations in terms of the UV graph and to be sure that findings do relate to that relationship between human-likeness and familiarity, where a dip should emerge where images are close to but not quite human. Secondly, stimulus materials should be carefully selected or developed to cover as wide a range of that human-likeness continuum as possible so it is possible to identify images representing defined points on that scale and to be able to say confidently that they are , for example, 25% or 75% human. Thirdly, the second study in this phase relied upon just five images to explore detailed questions about the nature of those faces so while the results can be used to suggest areas for future research, it is not possible to generalise the findings to all types of NHF as the patterns of description and response may actually be in response to those individual images. For example, without further research it is not possible to say with confidence that all robots would be described in the way found in this study because only one example robot image was used. A limitation of the study was that

it investigated just five face types in a lot of detail, and the anchor face would not be strong candidate for a UVE. To test whether the differences that this study hints at in terms of how faces are processed, a more systematic variation in human-likeness will be needed. Finally, a general issue should be acknowledged in that this study used a process of categorisation to allocate participants' comments to categories which were then aggregated to draw the broad conclusions on analytic versus holistic processing. It is acknowledged that this process builds on several assumptions between any processing that occurred when a participant viewed one of the stimuli and their behaviour in reporting their emotional responses and perceptions of willingness to share a home with the NHE.

These limitations were addressed in the design for the second and third research phases which developed more systematic and sophisticated methods for investigating the UVE.

In conclusion, when considering all of the results from this phase in terms of the UVE, it can be seen that the rating and description data has provided some limited support for the idea that an UVE exists. The hypothesised relationship between the rating variables were found, and while the correlational analyses do not allow causal conclusions to be drawn, they do serve as an indication that there is validity in Mori's broad concept of a relationship between human-likeness and strangeness/eeriness, and that some of the NHAs tested are able to reliably elicit an eerie and unsettled emotional response. However, this small-scale study was exploratory in nature with the intention of directing empirical studies to test possible explanations for why the UVE may occur and to understand the nature of uncanniness as a response. Certain decisions on scope and direction needed to be taken at this stage: computer graphics, motion capture and android design had progressed rapidly while this phase was in its design and analysis

stages, which greatly expanded the domain of near-human and virtual agents which would be available for investigation with new examples of embodied agents and high-quality avatars emerging on a regular basis. As well as offering a wider range of types of stimuli, this also presented a challenge as the original collection of potential UV-candidate stimuli used in this phase became out-dated. The body of literature looking into the UVE had also expanded during the period covered by this phase and several new studies looking at roles of mismatched expressions and anomalous facial features meant it was possible to identify parallels with the suggestions above that certain features were more salient than others in triggering the effect. This led to the decision to position the next two studies within the specific domain of face perception, synthesising the UV literature with existing literature how human faces are generally perceived, with the goal of understanding more about whether NHFs represented a different category of face-type stimuli and could be said to be being processed differently. The second phase of research developed the finding that participants used different approaches in describing the faces suggesting that there may be differences in the way that these faces are processed, and whether those processes for NHFs mirror those for natural human or artificial faces. The third and final research phase was designed to focus on developing the findings about different emotional responses to the faces by looking at facial expressions of emotion and drawing conclusions about the aspects of different faces or facial features that might contribute to a sense of the uncanny.

Chapter 4: Phase 2: The Inversion Effect

4.1: Introduction to Phase 2

This second research phase applies theories of face processing to the subject of the UVE by continuing to explore whether NHFs have eerie and unsettling qualities along with an examination of whether those NHFs are processed differently from how human or non-human faces are processed. The UVE is characterised as a particular pattern of emotional responses to NHAs as they vary in human-likeness, with a characteristic decline in positive emotion and increase in negative emotion at a specific point. The intention of the studies in Phase 2 was to begin to explore how this effect can be applied to the perception of faces, and in particular, whether it can be said that there is a relationship between how faces that vary in human-likeness are processed and how we respond to them. This is an important area for consideration as NHAs become more common-place, as their acceptability is likely to hinge on how people respond to them.

This phase explored the premise that NHFs are processed using an analytic approach rather than holistically, by looking for a presence of a FIE to test whether inverting NHFs would have the same detrimental effect on recognition performance that is found when human faces are inverted.

Section 1.4 above described how human faces are generally processed holistically, as an overall impression of the face rather than analytically in terms of their individual features. However, striking individual features were a notable characteristic of many of the NHFs explored in Study 1 where several had features that were particularly distinctive, especially as many had large or unusual eyes. This can be seen particularly striking in the eeriest face which was image 16, see Table 3 in Section 3.2.3. NHFs also represent a novel class of stimuli which may be processed in a unique manner. For example, robot faces

which are basically mechanical but possess some humanlike features and expressions may be encoded analytically, in a manner closer to objects, rather than holistically, as with human faces. This was explored in in Study 2 by looking at the differing patterns of holistic and analytic terms used when describing human, near-human and artificial faces. The Artificial, Cluster 1 and Cluster 3 faces were described either in analytic or holistic terms while the Cluster 2 and Human faces were described using a mix of both analytic and holistic terms. Analytic terms were used more when describing the Artificial, Cluster 1 and Cluster 2 faces, holistic terms more for this Cluster 3 and Human faces. This suggested an area for further enquiry to explore generally whether NHFs were processed analytically rather than holistically and to see if the pattern found for those five individual faces would be demonstrated in a wider range of face types.

The elicitation of a FIE has been used to explore holistic processing of faces. (Yin, 1969; Kanwisher and Moscovitch, 2009; Farah and Tanaka, 1995; Searcy and Bartlett, 1996, Freire et al, 2000). Inverting faces makes them harder to recognise, and the explanation for this has been that an upside-down face has a completely different configural appearance to the upright face. This means that it is not possible to use an encoded holistic representation to extract information about that face, making it harder to identity an individual or to know whether the face has been seen before. However, some factors can mitigate the magnitude of an inversion effect and Hills et al (2011) found that if participants were cued to pay particular attention to the eye region, the FIE decreased and those faces were easier to recognise compare to when no cues were given. Therefore, even when the configural information has been disrupted by inversion, an inverted face which has unusual or striking individual features may not be as detrimentally affected as a face without such distinctive features.

Chapter 1 described several studies which found an UVE for NHFs which have exaggerated or unusual eyes (Geller, (2008), Seyama and Nagayama (2007, 2009) and Looser and Wheatley, 2010.). If it is the case that NHFs are processed analytically rather than holistically, as a result of these unusual or anomalous features or for other reasons, there should be a decreased FIE for the NHFs compared to the human and non-human faces.

Applying the FIE to the UVE builds on the work of Seyama and Nagayama (2007, 2009) in applying face perception theory and techniques to the question of the UVE but takes a novel approach of examining the FIE as a measure of analytic versus holistic processing. This approach makes it possible to ask whether NHFs are processed more analytically than human or non-human faces. This question of whether analytic processing varies as a function of human-likeness was explored by two studies which applied the design principles outlined in the introduction to Phase 1 of this thesis. Having observed a feature that potentially contributes to an UVE and linked it to an area of psychological theory, the premise of analytic processing for NHFs was tested using stimuli that varied in human-likeness in a systematic and measurable fashion. The nature of the non-human end of the continuum was also varied by selecting different categories of images to allow an exploration of whether images taken from the 'classic' robot to human morph described in much of the UV literature would differ from images originating from other non-human agents such as statues, dolls and animals. Study 3 gathered subjective measurements of strangeness, eeriness and human-likeness, and then Study 4 used these rated images as experimental stimuli in a face discrimination task measuring recognition performance under different conditions of image orientation. These two studies will now be introduced in detail and the research questions they addressed will be presented.

4.2: Study 3: Ratings of morph faces for inversion study

4.2.1: Introduction to Ratings of morph faces

The main aim of this research phase was to answer the question of whether NHFs show evidence of greater analytical processing compared to human or non-human faces. This first study collected standardised data about the stimulus materials that were designed for that experiment. This introduction will present the types of images that were selected for inclusion in this study and recap on the measurements that were collected. Finally, the research question and hypotheses to be considered in this study will be presented.

As has been described above, the UVE describes a relationship between human-likeness and emotional response. In the majority of the UV literature, that human-likeness dimension is generally presented as a continuum beginning with a robot at the non-human starting point and ending with a real human at the opposite end. However, a robot is only one of several types of entities which exhibit human-like properties so a range of continua can be imagined which begin with other entities such as dolls or statues. It could be that the observations of eeriness that characterise the UVE are specific to the robot-human continuum and may not appear when other types of entity are placed at the other end of the continuum. While the main aim of this phase is to explore how NHFs are processed, a sub-goal addresses this question by using four different types of starting point for the non-human to human morphs. Robots were identified as the first category, as these are the classic example given by Mori (1970) as the entity that will evoke an UVE when humanised. Looser and Wheatley's (2010) research inspired the inclusion of dolls and statues as categories for exploration as, although they were not measuring eeriness and so weren't exploring the UVE directly, their work is relevant to the present study as it

explored a similar transition between non-human and human to find the point at which faces begin to appear lifelike. Finally, a fourth category of animals was added. The dolls, robots and statue faces shared the attribute that they were created with an intention to emulate human-like qualities, but the animal faces represented a class of face-like objects which do not bear an overt physical resemblance to human faces. This would allow an exploration of whether an UVE would emerge for this continuum but also whether there would be any evidence of analytic processing for these faces.

The desirability of collecting standardised ratings of eeriness, human-likeness and strangeness for any stimuli used to assess the UVE has been described in detail in the Introduction to this thesis. The rating scales that have been chosen for this thesis are those based on MacDorman (2006) as described in Phase 1.

Research Question

How will the subjective ratings of human-likeness, strangeness and eeriness compare for a collection of morphed face images created as animals, dolls, robots and statues are progressively morphed into matched humans? This research question mirrors those posed in Study 1 and Study 2, and is the basic test of whether there is evidence for an UVE. As such, it is based mainly on Mori's proposal of the UVE and the shape of his hypothesised graph where familiarity gradually increases with human-likeness until a valley is found at the almost but not quite human point.

H1. The measured human-likeness for each face will increase with each morph stage, with the non-human end point rated as the least human and the final human image rated as most human. This may seem like a common-sense prediction but if it is supported it will provide evidence that the morph technique does indeed represent a way to create a gradual progression in human-likeness against which the eeriness and strangeness ratings could be reasonably compared.

H2. Ratings of strangeness will be significantly higher for the near-human morphs (75% human/25% non-human).

H3. Ratings of eeriness will be significantly higher for the near-human morphs (75% human/25% non-human).

The UVE is characterised by a peak of both strangeness and eeriness at the near-human point: support for these two hypotheses will provide evidence that a UVE does occur with these morphed faces.

In interpreting the results of these hypotheses, as well as looking at the overall ratings of eeriness, strangeness and human-likeness, the results will be examined by category to whether the UVE-indicative pattern will be found for each of the four image categories: animals, dolls, robots and statues. As the UVE was initially proposed as an artefact of making robots more human-like, it is a reasonable prediction that the robot category will show this pattern. The doll and statue categories were added based on the work of Looser and Wheatley, (2010) who found that perceptions of animacy emerged at the near-human point. They theorised that awareness of whether something is alive or not may contribute to the UVE and so it is predicted that a valley pattern will emerge for these two categories as well. However, the animal category has been added as a novel exploration to see if it does demonstrate an UVE so no specific prediction of the direction has been made.

4.2.2: Method

Design

A 4 x 5 repeated measures design was used to explore the factors of image category and morph stage. There were four levels of image category: animals, dolls, robots and statues and five levels of morph stage: non-human, 25% human, 50% human, 75% human and human. Three dependent variables were measured, with participants rating 60 images on the following three scales:

Human-likeness: 9 point scale from 'very mechanical' to 'very human'

Strangeness: 10 point scale from 'very familiar' to 'very strange'

Eeriness: 10 points from 'not eerie' to 'extremely eerie'.

Participants provided 180 ratings each, one each of three ratings for each of the 60 images. Presentation order was randomised to control for order effects.

Materials - Stimulus Images

Sixty photographic quality images were produced from twelve base pairs of human and non-human photographs of faces, by morphing the non-human images into the human images. Stills were taken at the 25%, 50% and 75% human point, which produced five morph stage images for each base pair. Four categories of non-human initial images were used (animals, dolls, robots and statues) with three examples in each category. The five 'morph stages' were the original non-human and human photographs plus intermediate still images taken from the morphing process at the points where the morphed face was 25% human/75% non-human, 50% human/50% non-human, 75% human/25% non-human. An example of a base pair plus the three intermediate images is shown in the figure below:

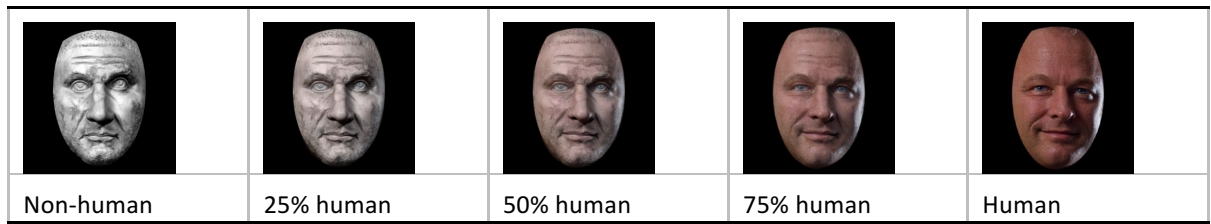


Figure 15: Image set for one example of a statue to human morph.

An example illustrating each category is shown in Figure 16 below.

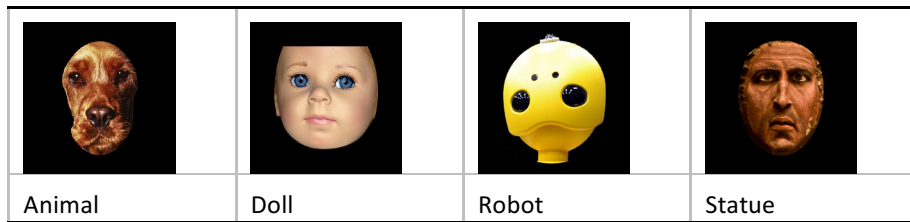


Figure 16: Image set for one example from each image category.

Each set of five morph stage images was produced using the same method. Firstly, an internet search was carried out to find three starting images in each category of non-human image plus a matched human face. These pairs of images were used as the start and end points to create an animation systematically transforming the starting face into the end face. Finally, still images were taken from the animation at defined points so these stills could represent the desired five points along the human-likeness continuum. Each of the stages in this method will now be described in detail.

Collecting source images

Following the theme of this research project to date, it was decided that the first criterion should be that they were faces that participants may have encountered in the course of their day to day life rather than ones manufactured to be deliberately unsettling. To this end, a general internet search for existing photographs was chosen over the creation of posed new images of faces for each category. Criteria were set to narrow the search to appropriate images and to ensure consistent presentation:

- The photographs had to be a clear and unambiguous representation of the category; for example, even very realistic statues needed to clearly portray that they were crafted from an artificial material.
- The pictures had to show the full face presented frontally and had to avoid having any part of the face obscured by hair, shadow or spectacles.
- Start and end point images had to match facial expression as closely as possible in each pair.
- The images had to be high quality as the morph process works best with images that are crisp and clear at their original resolution.
- Only images which stated that they were produced as royalty-free and available for general use or ones that had been shared with the appropriate creative commons permissions to allow research use with acknowledgement.

Appendix 2c presents a list of each of the source images with their original web link.

Issues considered in the selection of the corresponding human images

As well as ensuring that the corresponding human faces showed a comparable expression to their non-human pair, human faces were sought which were matched as closely as possible in terms of colouring, shape and size. This was easiest for the doll and statue images but harder to achieve for the robot and animal images. (Two of the animal images were sourced from pet food commercials where ‘owners’ and their pets were presented side by side to emphasise the similarities between their appearances. a humorous advertising technique that proved to be very useful for the matching quandary!) In some instances it was impossible to find a suitable human match which met the above quality criteria and as a result that non-human face was discarded. The final set of images therefore represented non-human faces which could be matched to a freely available human image of suitable quality for the morphing process.

Cleansing of images

The source images presented the human and non-human faces against a range of different backgrounds. The morph process required the images to be as 'clean' as possible with either a standardised identical background or no background at all. To accommodate this, the source photographs were masked so that the face area was presented against a plain black background and any extraneous items were removed. Where there were small discrepancies between the sizes of the two images these were also corrected at this stage.

Designing morph sequences

The chosen morphing software (Abrasoftware Fantamorph SE) produces short animated sequences where a starting image is transformed frame by frame into a different image. This process gradually blends the contents of one image into another using shared key points on the two images. (It is possible to combine more than two images but for the purposes of this research only pairs of images were used.) When a key point is selected on the first image, a corresponding point is automatically highlighted on the second image and the user can either accept its positioning or adjust it to move it closer to the correct position. Individual points can be used to produce simple morphs but a more powerful technique is to use several points in the same area connected by lines to define the boundaries of individual features such as the eyes or to indicate where there is a region on one face that does not correspond exactly to the other face in the pair. The figure below illustrates this principle:



Figure 17: Screenshot of the Fantamorph software package, showing the set-up of correspondence dots for two faces.

The key points are used by the software to produce an automated animation between the two images. Where images are similar this usually has straightforward and predictable results but trying to morph between features with very different appearances can occasionally have unexpected consequences. This can result in the production of unwanted 'ghosting' where, for example, whiskers on an animal face appear as a prominent feature of the later human stages of the morph or 'tearing' where the software cannot accommodate large differences in appearance and blacks out areas on some frames. The only solution to these problems is to carefully review the placement of the key points to mask the unwanted features at that stage. While it is possible to create face templates which can be re-used for different images to speed up the morph design process that was not a useful shortcut for this project as the non-human images varied widely in the arrangement of the features. Designing the morph sequence for each pair

required careful review of the individual features, selection of appropriate key points and adjustment and tuning of the animated sequences.

Exporting final images

Once the animation has been produced and judged to be a satisfactory morph between the two images, the final stage is to export the frames at the desired defined points. Regardless of the size of the starting images, all stills were exported at a size of 400 x 400 pixels.

Materials - Study Web Pages

Web Survey Pages

An online experiment was used to present the sixty images and collect participant ratings on each of the three scales. The web survey was produced using the same page templates as those used in Study 2 described above.

Pilot

Two volunteers offered to test the survey and give critical feedback to help improve the final product. While there was no technical problem in working through the web pages, they felt that that rating 60 images on three different scales - a total of 180 images to look at and click through - was an arduous task that risked participants dropping out part of the way through. To address this, consideration was given to the best method to allow fast ratings to be provided while ensuring that participants gave accurate answers. In the version presented to the pilot testers, the same method was used for the rating Study 1 where participants chose their answer on the scale and then clicked a button to submit it and move to the next image. This effectively doubled the number of clicks that a participant needed to make; one for the rating selection and one to submit it and see the next image. For the earlier study where only fifteen images were being rated this was tolerable but for a larger number of images it became time-consuming. The code was

amended so that the page would advance automatically as soon as a response button in the scale was clicked. If participants did accidentally click on the 'submit this response' button without a making a selection on the scale, another image/scale combination was presented but the counter did not increment and the combination that had been seen but not answered went back into the pool of responses still to be completed. This version was reviewed by the testers who commented that it was an improvement and allowed them to complete the survey in a shorter time.

Results Database

The study results (ratings and demographics) were again stored in a Lotus Notes database hosted at the Open University, built using the same design principles and data security considerations as those previously described. Only fully completed responses were passed through to the database as any participants who abandoned the study part of the way through were deemed to have withdrawn and their partial responses were not stored.

Procedure

As with the previous web study, participants were recruited from a variety of routes. social media links were used as before, as well as direct recruitment of participants who had expressed an interest in the research. In addition, three new sources of potential participants were used as advertisements were placed on the PsyPAG mailing list used by post-graduate psychology researchers and two web sites³ that exist to promote psychology research studies to interested parties. All participants arrived at the online study by clicking on a web link. The structure of the study was identical for everyone but

³ <http://psych.hanover.edu/research/exponnet.html> and

<http://www.onlinepsychresearch.co.uk/researchers>

the 180 pages were randomly ordered for each participant, with the constraint that it was not possible to see the same image twice in a row. Example screenshots of the final web pages used in study are included in Appendix 2e. The structure of the experiment was as follows:

Introduction: The initial landing page introduced the study and the researcher and gave background information about the benefits and implications of participation. Participants could only proceed to take part in the study if they clicked to indicate that they were over eighteen years old and had understood the terms and conditions sufficiently to give their consent.

Task screens: 180 task screens were then presented one by one. Each page displayed one of the sixty images to be rated in the top part of the screen and one of the three scales beneath the image. The number of images that had been rated, the number that were still to come and this ratio expressed as a percentage were all presented next to the image to indicate progress through the study.

Participants made a selection rating the image on the scale and then the next page was presented.

Demographic information: When all of the images had been presented and rated, participants were asked if they were willing to give some basic demographic information. These optional questions (gender, age and location) were the same as those collected in the first web study described above. Finally, participants were asked if they would be willing to take part in future experiments and if so, to provide their preferred email address.

Close: Once the final questions had been answered, participants were routed to the main project web page for further reading about the UVT. The landing page repeated the contact information for the researcher if there were any further queries.

Screenshots showing an example of each page in the structure are given in Appendix

2e.

Participants

107 participants completed Study 3. 105 participants provided demographic information; 68% of those were female and the mean age was 38. When asked about their location, most participants (80%) reported that their location was in the United Kingdom. A full breakdown of the demographic data is shown in Appendix 2a.

4.2.3: Results

H1. The measured human-likeness for each face will increase with each morph stage, with the non-human end point rated as the least human and the final human image rated as most human.

Participants rated all sixty faces on all three scales. Firstly, the ratings of **human-likeness** will be described and analysed. Human-likeness was measured on a 9 point scale, 9 where 1 was labelled as 'very mechanical' and 9 'very human'.

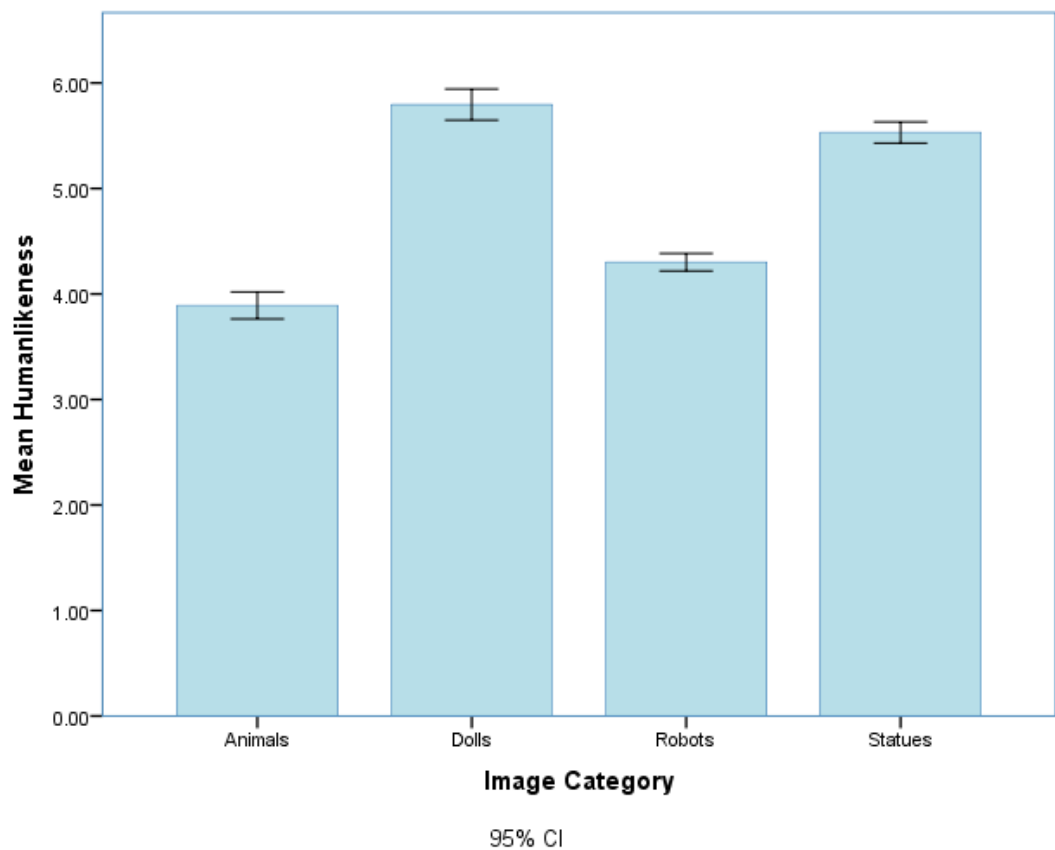


Figure 18: Mean human-likeness with error bars for each image category.

The bar chart above presents the mean human-likeness ratings for each image category. They are based on figures derived by collapsing rating scores across all five morph categories and all three individual images in each category to give an overall mean human-likeness rating. At the overall category level, none were rated as highly human-like but the doll images were rated as the most human-like ($M=5.8$) followed by the statues ($M=5.5$). The robots were the next human-like ($M=4.3$) with the animals rated as least human-like of all ($M=3.9$).

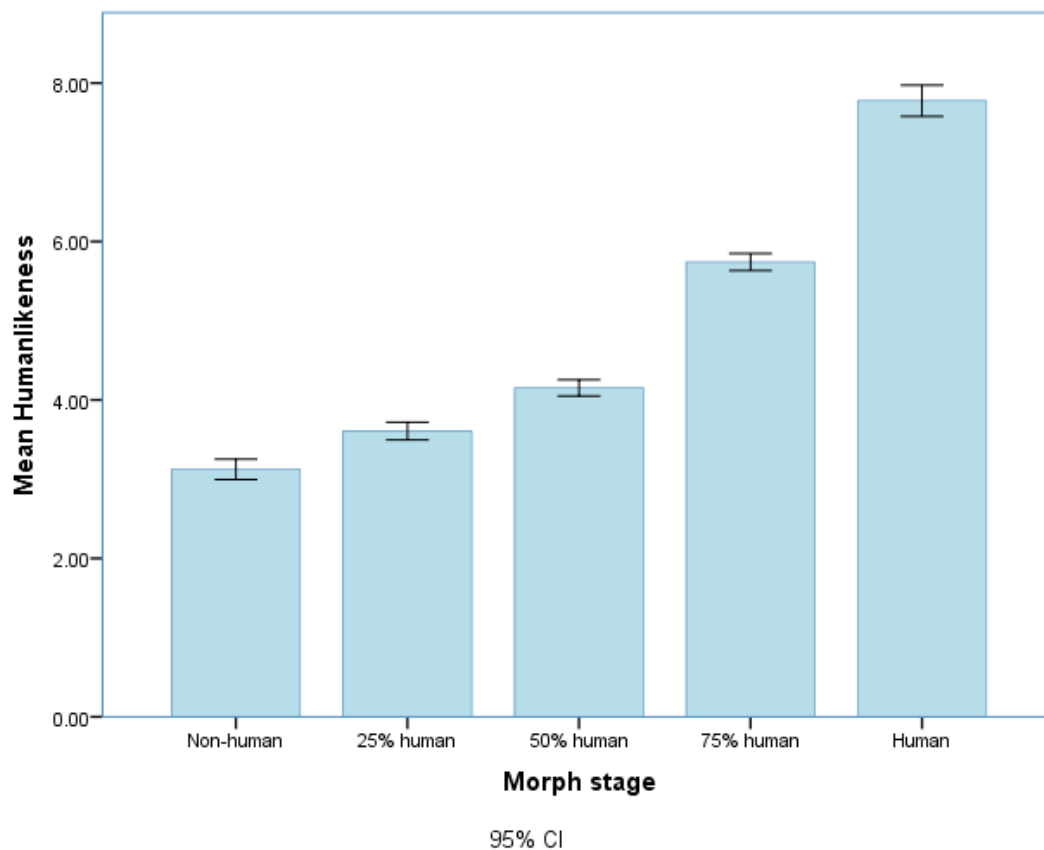


Figure 19: Mean human-likeness with error bars for each morph stage.

Figure 19 above presents the mean human-likeness for each morph stage. The mean ratings show a clear progression of increasing human-likeness as the morph stages progress from non-human to human and suggests that the morph stages were perceived, as they were designed, to represent a continuum of human-likeness for the different image categories. However, although the morph images were taken from the morph animation at equidistant intervals, the ratings do not show such a linear progression. It could be expected that the magnitude of these differences in human-likeness rating would also be equal but this was not the case as is shown in Table 10 below.

| Morph Stage | Difference between means |
|------------------|--------------------------|
| Non-human to 25% | 0.5 |
| 25% to 50% | 0.5 |
| 50% to 75% | 1.6 |
| 75% to Human | 2.0 |

Table 10: Increases in human-likeness between morph stages.

It can be seen that the differences between the categories increase as the morphs progress from less human to more human.

When the rating results were analysed using a 4 x 5 repeated measures ANOVA, Mauchley's test indicated that the assumption of sphericity had been violated for the main effects of category and morph stage, as well as for the interaction between them:

Category: $\chi^2(5)=55.23, p < .0001, \epsilon=.76$.

Morph stage: $\chi^2(9)=198.86, p < .0001, \epsilon=.48$.

Interaction: $\chi^2(77)=178.28, p < .0001, \epsilon=.76$.

When interpreting the ANOVA results the degrees of freedom were corrected to accommodate this using the Greenhouse-Geisser estimates of sphericity for the morph stage and Huynh-Feldt estimates for the category and the interaction.

A significant interaction between image category and morph stage was found. ($F(10.05,1065.57)=63.75, p < .0001, \text{partial } \eta^2 = .38$) In addition to the significant interaction between the image category and morph stage, there was a significant main effect of the category of the non-human image, $F(2.27,240.36)=182.98, p < .0001, \text{partial } \eta^2 = .63$) and the main effect of morph stage was also significant, $F(1.92,203.41)=626.98, p < .0001, \text{partial } \eta^2 = .86$). Considering these findings in the light of the hypothesis detailed above, the presence of a significant interaction indicates that the image category had different effects on the ratings of human-likeness at each morph stage. A visualisation of this interaction is shown in the figure below, showing that ratings of human-likeness increase for all image categories as the morph stage progresses from non-human to human, but that the pattern of this increase varies according to the type of initial image.

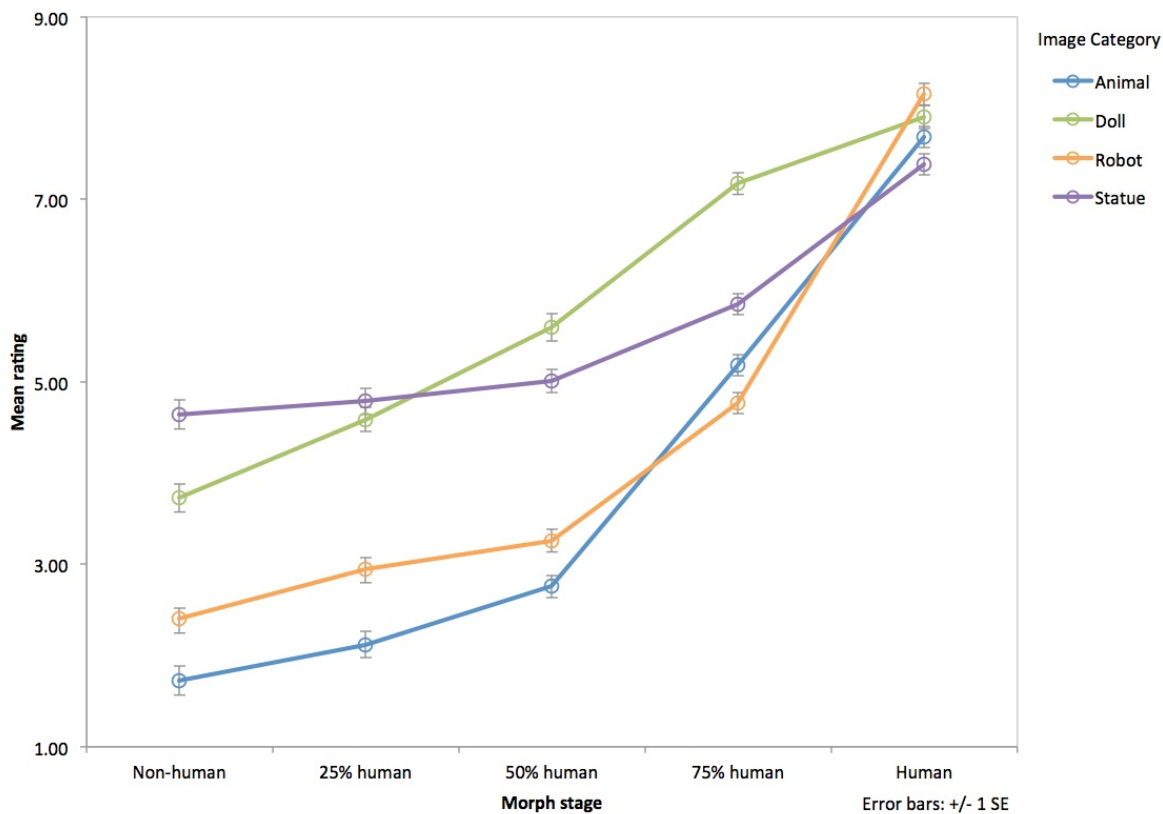


Figure 20: A comparison of human-likeness ratings for four image categories as the morphed faces progressed from non-human to human

The robot category is the most useful reference point in reviewing the variation by category as it is the classic example from Mori's proposal of an UVE and should show a steady increase in ratings of human-likeness from non-human to human. The robot morphs do show the expected progression with increases in human-likeness across the morph stages but this is not a continuous progression, with steeper increases between the 50% to 75% human stages, and the 75% to human stage. The novel animal category showed a similar pattern to the robots, as both had low ratings of human-likeness at the non-human, 25% and 50% stages, but the animals were rated as slightly more humanlike than the robots at the 75% stage, and slightly less at the human stage. The dolls were rated as more human in their un-morphed state than the robots and animals, which makes sense considering that they are objects designed to look like human infants. Their human-likeness ratings increased with each morph stage as predicted in H1. Finally, the

statues had the highest ratings for human-likeness in their original, un-morphed state and the lowest ratings for human-likeness at the fully human stage. The human-likeness ratings did increase with each morph stage but the pattern was markedly different from the other categories, with a much shallower curve showing only very small increases between the non-human and 25%, and 25% and human stages. The high initial ratings may be explained because statues are objects which are designed to look like people so began at a higher level of human-likeness than any of the other image categories.

H2. Ratings of strangeness will be significantly higher for the near-human morphs (75% human/25% non-human).

Participant ratings of the **strangeness** of the images will now be described and analysed. Strangeness was measured on a 10 point scale with the low anchor labelled 'very familiar' and the high anchor 'very strange'.

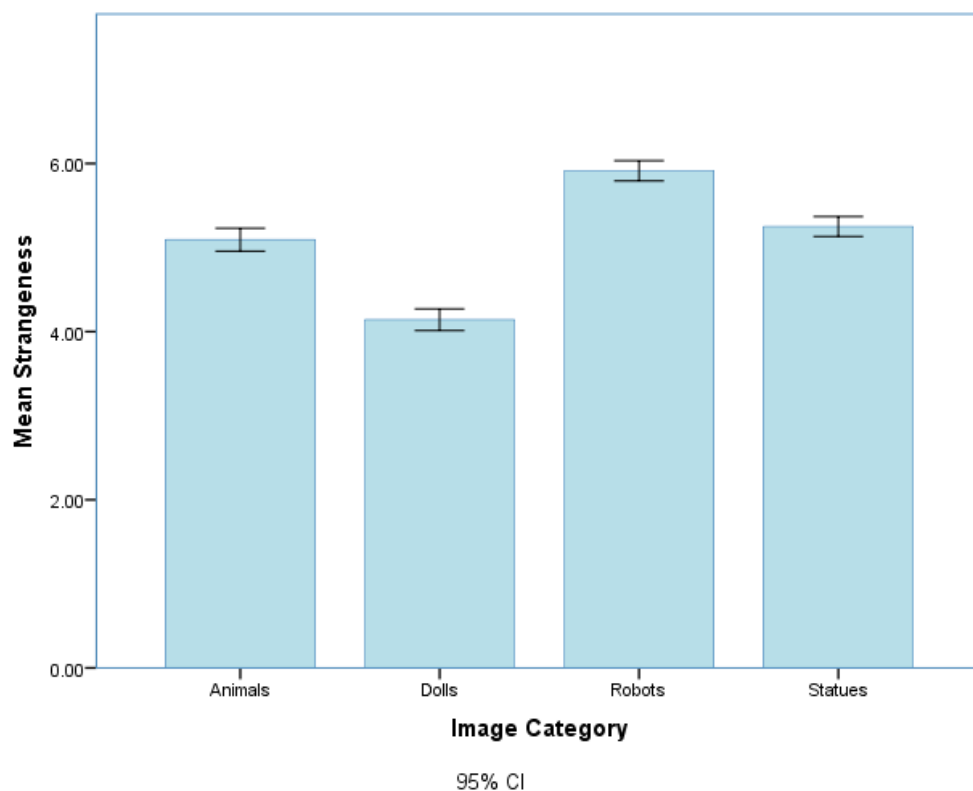


Figure 21: Mean strangeness with error bars for each image category.

The strangest faces (collapsed across the three example images and the five morph stages) were the robots ($M=5.9$) and *statues* ($M=5.2$) with the animals the next strange ($M=5.0$) and the *dolls* the least strange of all ($M=4.1$). As noted above, robots were initially proposed as the non-human entities that display an UVE as they become more human-like and as such serve as a reference category for comparison to the other three types.

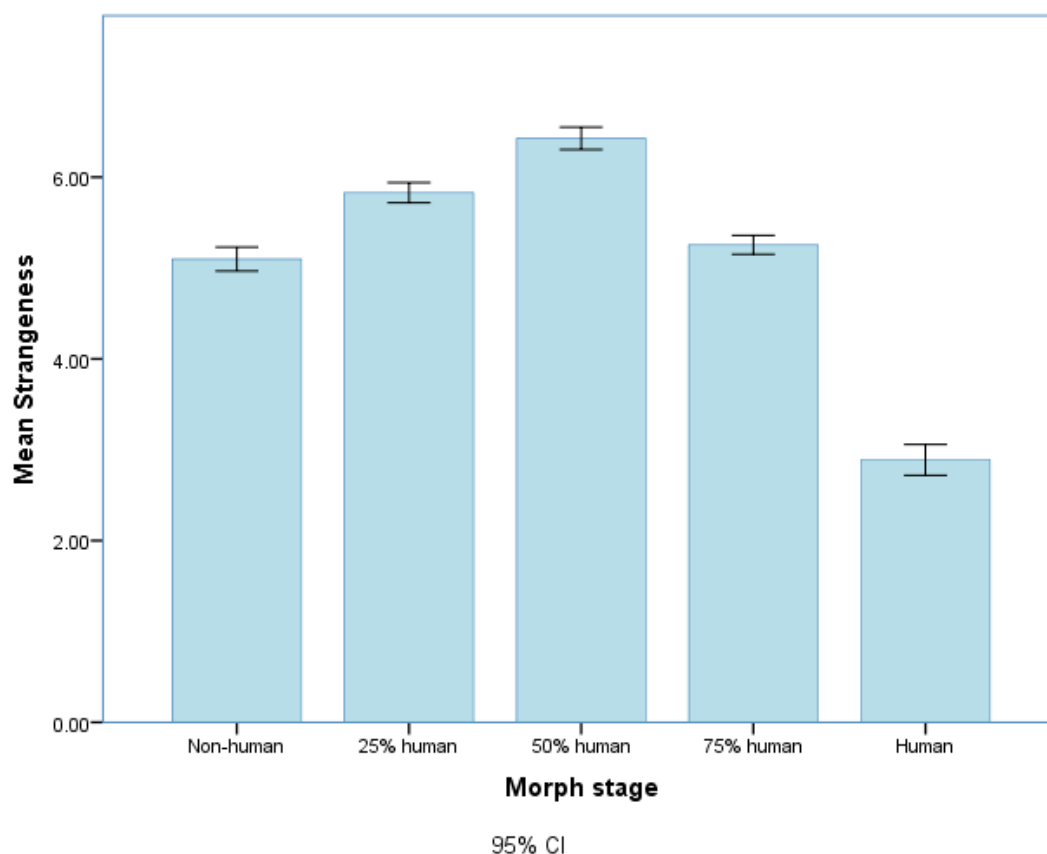


Figure 22: Mean strangeness with error bars for each morph stage.

The strangeness ratings for the five morph categories (collapsed across the four categories and three images within each category) show a pattern of an initial rise from non-human to 25%, peaking with the strangest face being the 50% human ($M=6.4$) before decreasing again with the humans rated as least strange, ($M=2.9$) as would be expected.

When the strangeness ratings were analysed using a 4 x 5 repeated measures ANOVA, Mauchley's test indicated that the assumption of sphericity had been violated for the

main effect of the morph stage ($\chi^2(9)=101.75, p < .0001, \epsilon=.70$) and for the interaction between the image category and morph stage $\chi^2(77)=223.22, p < .0001, \epsilon=.70$). (This was not the case for the main effect of category where $p = .62$. When interpreting the ANOVA results the degrees of freedom were corrected for the interaction and main effect of morph stage using the Greenhouse-Geisser estimates of sphericity.

All effects were found to be significant at $p < .05$: the interaction between image category and morph stage was found to be significant, $F(8.15, 863.40)=118.89, p < .0001$, partial $\eta^2 = .53$, as well as a significant main effect of the image category, $F(3, 318)=98.73, p < .0001$, partial $\eta^2 = .48$, and of the morph stage, $F(2.81, 298.13)=334.56, p < .0001$, partial $\eta^2 = .76$. The significant interaction indicates that although strangeness does indeed vary by morph stage and category, the pattern of the differences in strangeness between different the morph stages varies according to the type of image that was used as the non-human endpoint of the morph.

A visualisation of how the interaction between morph stage and image category impacted on strangeness ratings is shown in Figure 23 below:

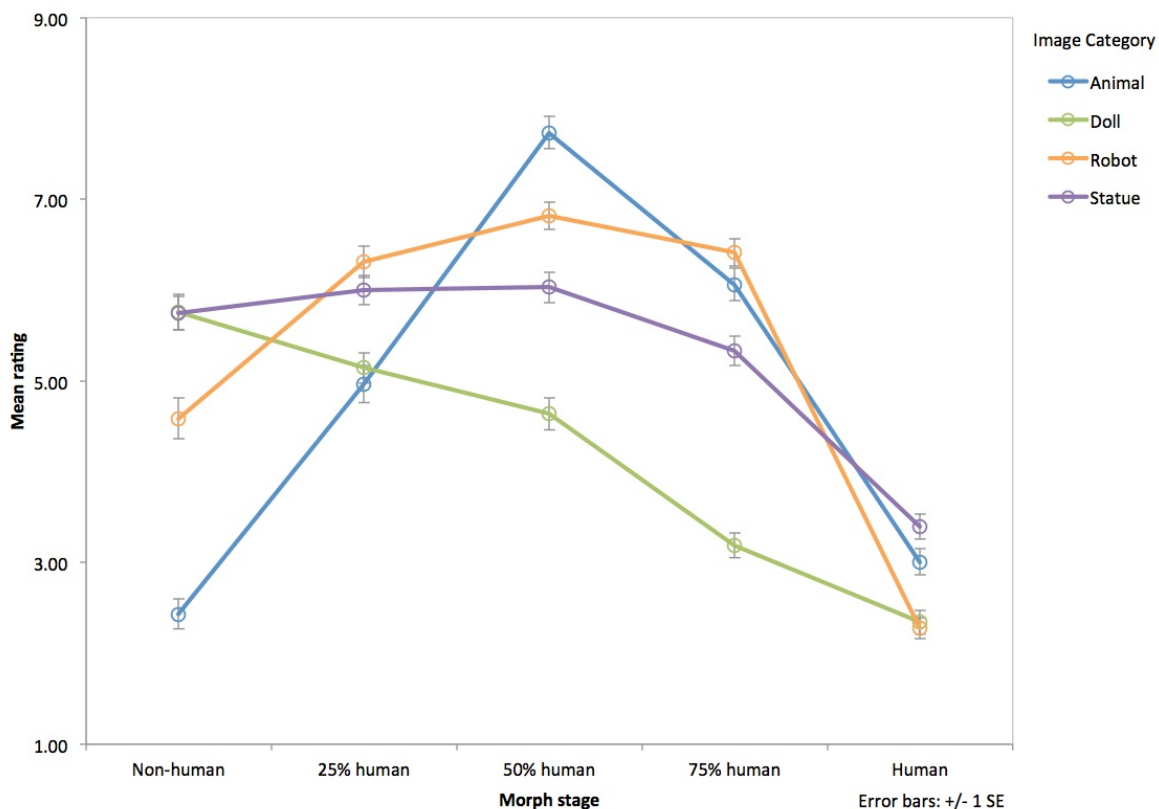


Figure 23: A comparison of strangeness ratings for four image categories as the morphed faces progressed from non-human to human

H2 predicted that ratings of strangeness would be highest for the 75% morph but this was not actually found for any of the categories. Following Mori's graph, there should have been a gradual increase in strangeness up to the 75% morph stage, after which it should have decreased sharply with the human faces being rated as not particularly strange. Considering robots first, these were rated strangest of all four categories in their un-morphed state, and strangest robot morphs were actually those at the 25% and 50% points. The mean ratings at those stages were very similar with $M=7.115$ for the 25% stage, and 7.121 at the 50% stage. Strangeness sharply decreased from the 50% stage to the human stage. The strangeness ratings for statues followed a very similar pattern to those for robots, albeit at a lower level of strangeness across each of the morph stages. This may be explained as the initial statues were designed to look like people while the robots were designed to approximate human-likeness to varying degrees, but still

retained some aspects of an artificial appearance. The animal category did have a high peak of strangeness, rated as the strangest of all of the morphs, with a steep increase in strangeness between non-human and 50% and a correspondingly steep decrease in strangeness from the 50% stage to the human stage. The doll category is particularly interesting here as it is the only one to show a consistent decrease in strangeness across the morph stages. Dolls have been proposed as a type of NHA which may be seen as inherently eerie so it is notable here that their strangeness ratings decreased at each morph stage.

H3. Ratings of eeriness will be significantly higher for the near-human morphs (75% human/25% non-human).

Finally, the ratings of **eeriness** will now be described and analysed.

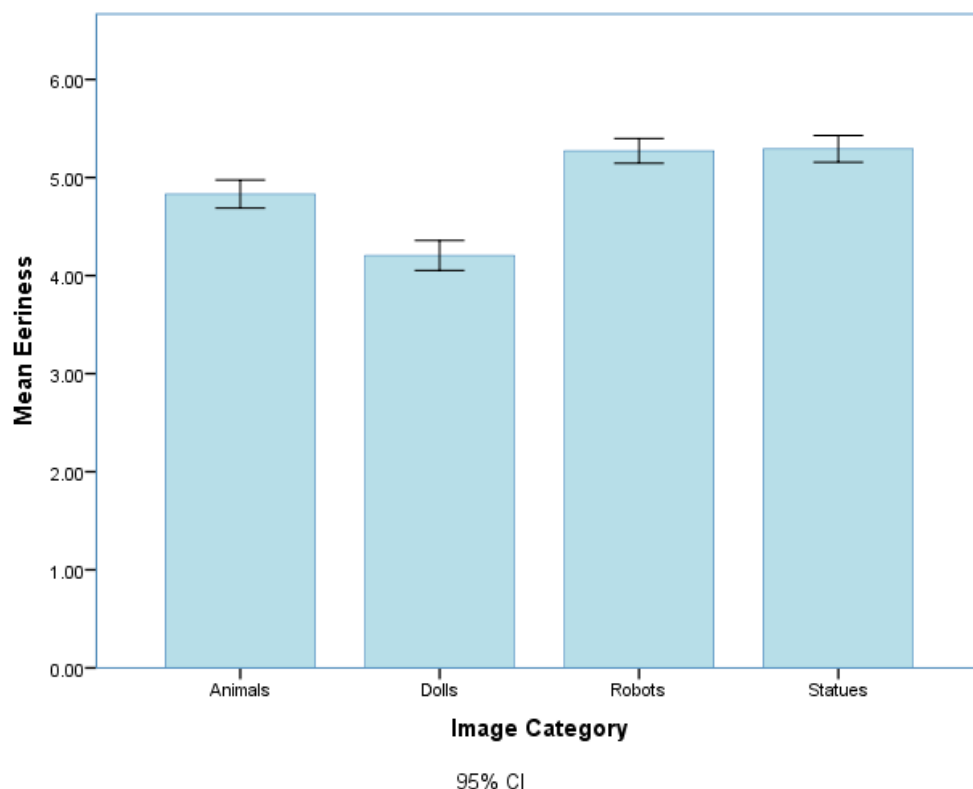


Figure 24: Mean eeriness with standard error bars for each image category.

Reviewing the mean eeriness by category (collapsed across morph stage and individual image) the eeriest categories were the robot and statue images with similar

mean eeriness ratings of $M=5.27$ and $M=5.29$. The animal category was slightly less eerie with a mean rating of 4.8 but the least eerie was the doll category with a mean rating of 4.2.

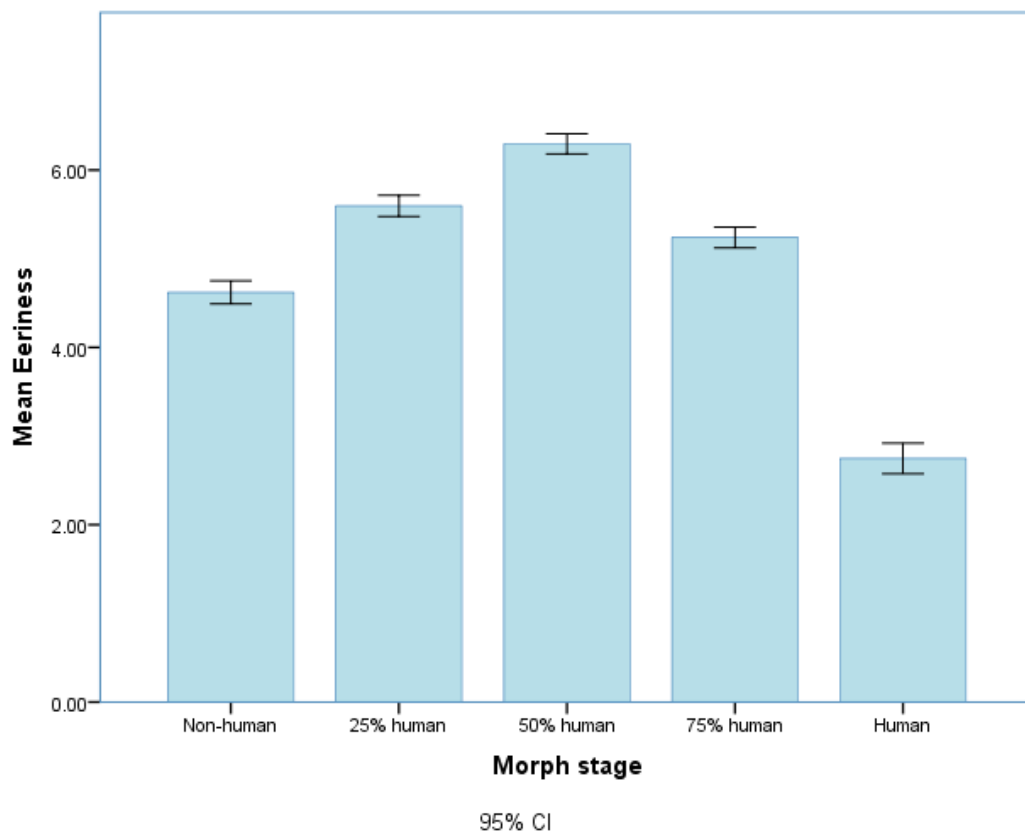


Figure 25: Mean eeriness with error bars for each morph stage.

The mean eeriness ratings for each morph stage (collapsed across category and individual image) followed the same type of pattern as the strangeness ratings previously described, with the 50% human morph stage rated as the eeriest ($M=6.3$), with the human ($M=2.75$) and non-human ($M=4.6$) stages rated as the least eerie.

When the rating results were analysed using a 4 x 5 repeated measures ANOVA, Mauchley's test indicated that the assumption of sphericity had been violated for the main effect of morph stage, $\chi^2(9)=57.6$, $p < .0001$ and the interaction between category and morph stage, $\chi^2(77)=250.2$, $p < .0001$. The assumption of sphericity for the main effect of category was not violated. To accommodate the above violations, the degrees of

freedom were corrected using the Greenhouse-Geisser estimates of sphericity using $\epsilon=.77$ for the main effect of morph stage and $\epsilon=.67$ for the interaction.

All effects were found to be significant at $p < .05$. There was a significant interaction effect between the image category and the morph stage, $F(8.1,857.8)=106.59$, $p < .0001$, partial $\eta^2 = .50$. The main effects of the category of the non-human image and the morph category were also significant: category = $F(3,318)=38.95$, $p < .0001$, partial $\eta^2 = .76$. Morph stage = $F(3.83,327.2)=328.45$, $p < .0001$, partial $\eta^2 = .76$. The significant interaction demonstrates that the image category had different effects on the ratings of eeriness at each of the morph stages. A visualisation of this interaction is shown in the figure below:

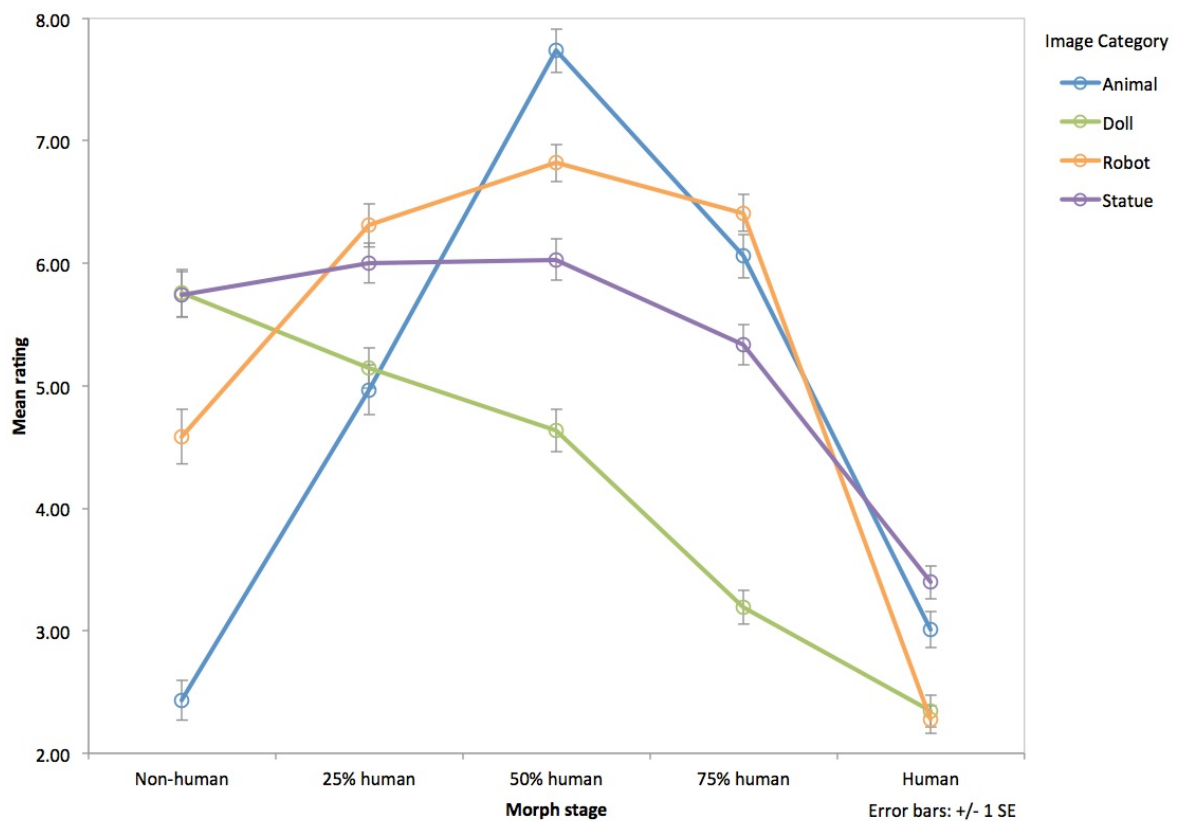


Figure 26: A comparison of eeriness ratings for four image categories as the morphed faces progressed from non-human to human

H3 proposed that the 75% human category would be rated as significantly more eerie than the others but this was not supported, as none of the categories were rated as most eerie at the 75% human stage. However, the pattern of eeriness ratings across the five

morph stages varied significantly by category and so these will be considered separately. The robot images were rated as moderately strange in their un-morphed state, and increased in eeriness with the highest ratings of eeriness at the 25%, 50% and 75% stages.

It was decided to focus in detail on the differences between these stages and explore whether the differences between pairs of mean ratings of eeriness were significant. To do this, two repeated measures ANOVAs were carried out, with one looking at the differences in eeriness for the robots at the 25% to 50% stages and the other looking at the 50% to 75% stages, and these were both found to be statistically significant. For 25% stage to 50% stage, $F(1,106)=16.18$, $p < .0001$, partial $\eta^2 = .132$, and for the 50% to 75% stages $F(1,106)=8.77$, $p < .0001$, partial $\eta^2 = .076$. It can be concluded that the level of human-likeness does indeed have a significant effect on how unsettling participants found the images, and that the changes between the morph stages representing low, medium and high levels of human-likeness are significant.

The eeriness ratings for statues again started high, but only increased modestly from the non-human to 25%, then 25% to 50% stages, before decreasing sharply from 50% to 75% and again from 75% to human.

The eeriness ratings for the animal category mirrored their strangeness ratings in starting with the lowest eeriness ratings of all four categories, being peaking with the highest ratings of eeriness at the 50% human point and then sharply decreasing in eeriness between the 50% and 75% stages, and the 75% and human stages. The eeriness ratings of dolls followed a unique pattern in these four categories by decreasing in eeriness at each stage, a similar pattern to that found for the strangeness ratings, which may provide some evidence that people find dolls intrinsically strange and eerie and the act of morphing them to become more humanlike removes some of those unsettling properties.

Visualising human-likeness and strangeness

Mori's UVE idea relies upon the relationship between human-likeness and strangeness, so in addition to the hypotheses based on specific measures, the mean values of human-likeness and strangeness were plotted against each other for each image category, with a predicted curve was been overlaid on these charts to shows the relationship that would be expected in the UVE. As in Study 2 above, note that this curve shows a peak rather than a valley as the ratings in this study measured strangeness rather than familiarity. None of the image categories showed the predicted relationship between human-likeness and strangeness.

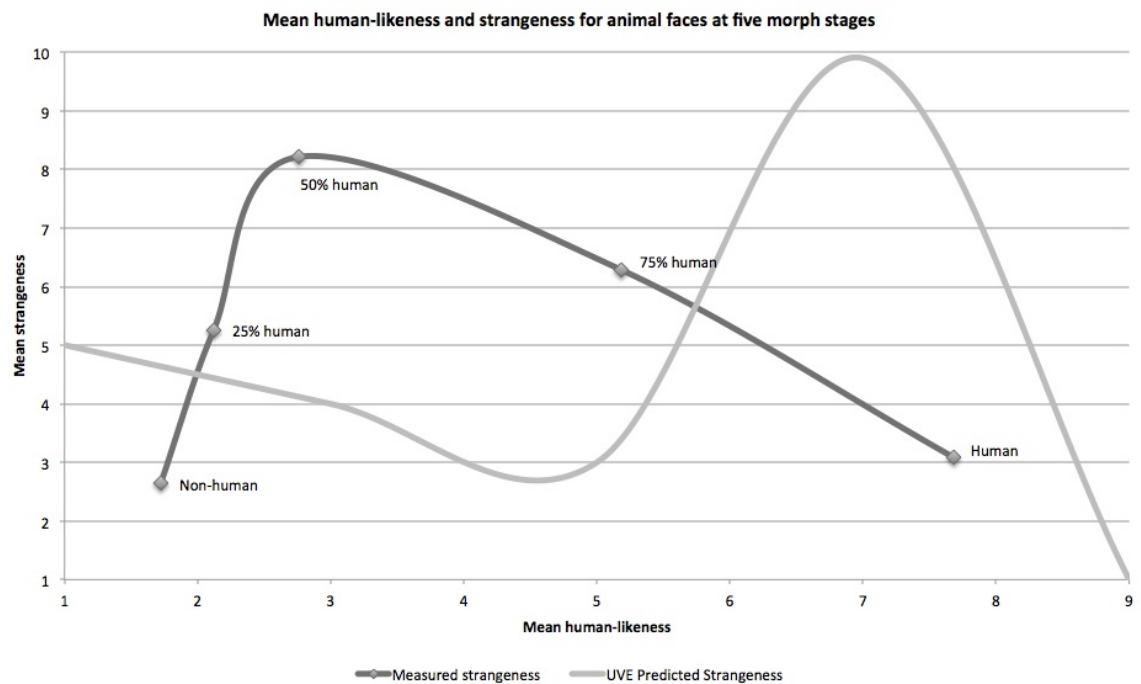


Figure 27: Mean human-likeness and strangeness for animal faces

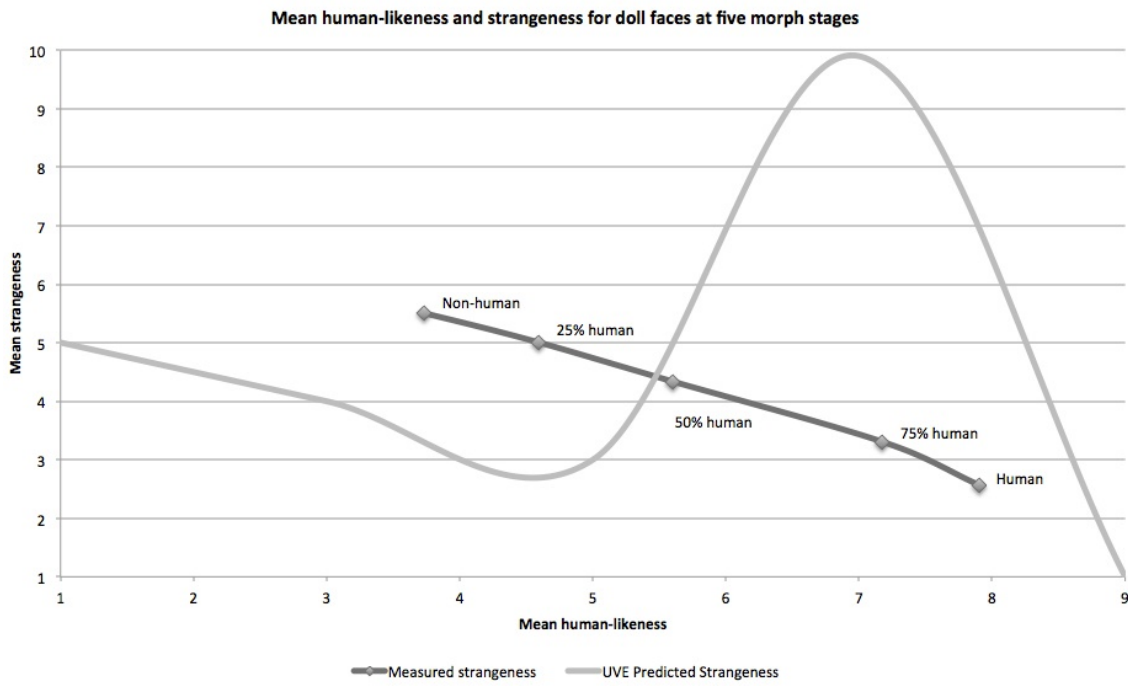


Figure 28: Mean human-likeness and strangeness for doll faces

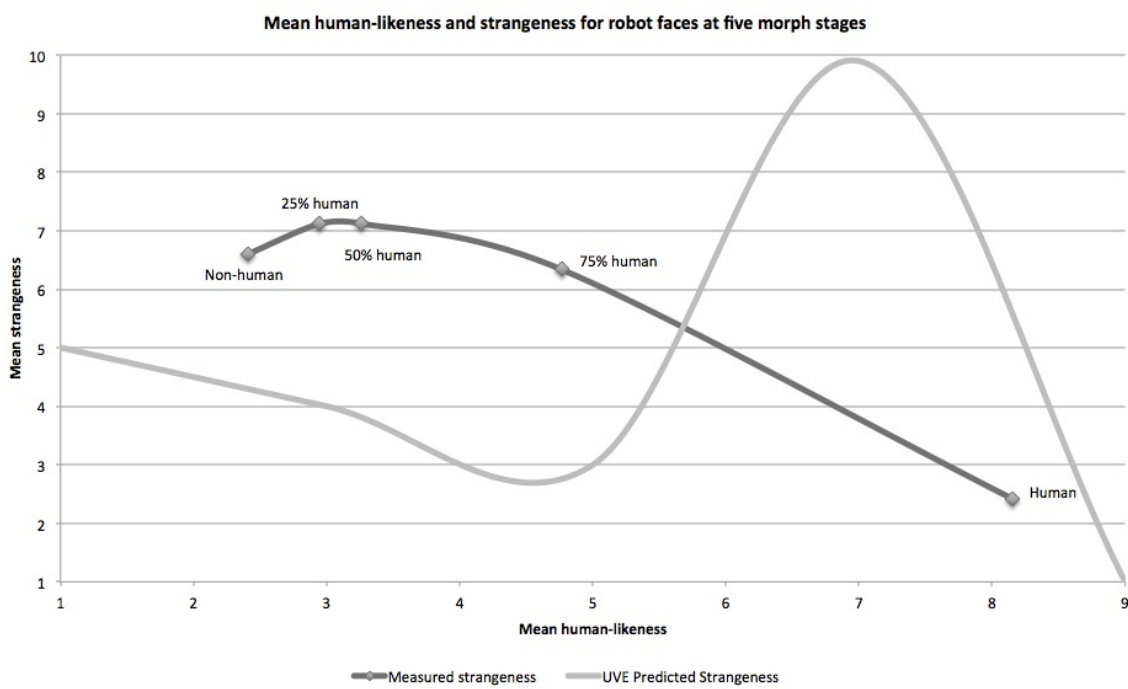


Figure 29: Mean human-likeness and strangeness for robot faces

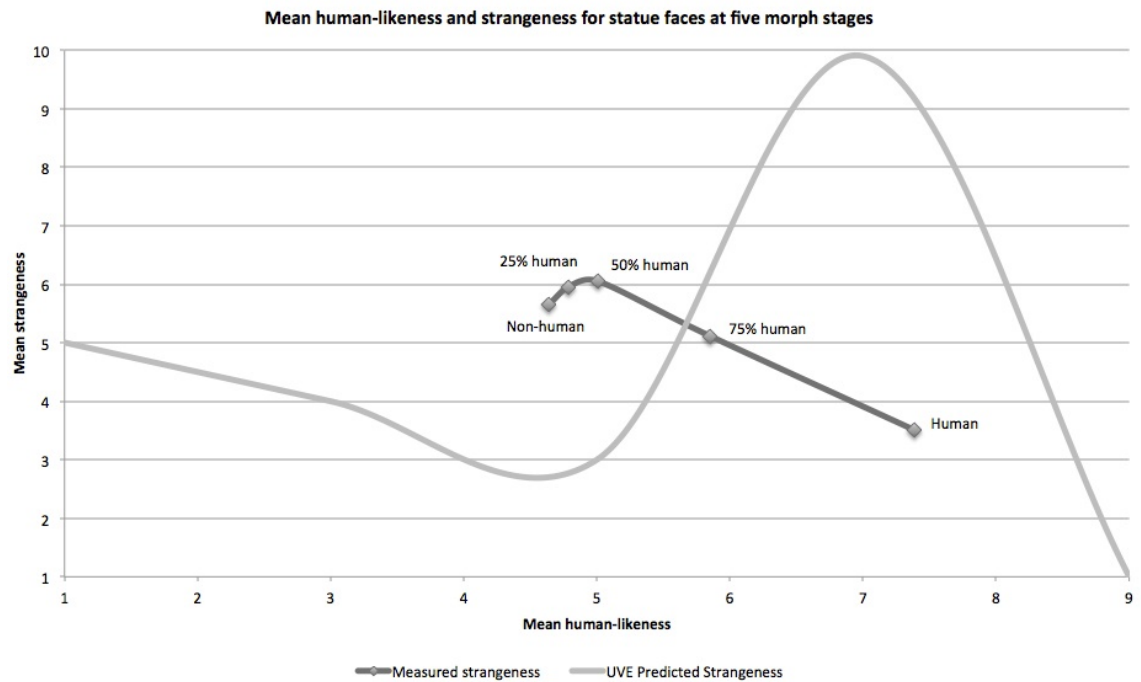


Figure 30: Mean human-likeness and strangeness for statue faces

4.2.4: Discussion

How will the subjective ratings of human-likeness, strangeness and eeriness compare for a collection of face images created by progressively morphing images of animals, dolls, robots and statues into matched human faces? These measures test for evidence of an UVE, based on Mori's hypothesised graph where familiarity gradually increases with human-likeness until a valley is found at the almost but not quite human point. Study 3 tested three hypotheses which predicted a steady increase in ratings of human-likeness (H1) and peaks in eeriness and strangeness for the morphs at the near-human point (H2 and H3). This study was designed to examine whether those ratings were affected by the type of image used to start the morph, looking at four categories of animals, dolls, robots and statues. The UVE was originally proposed as the result of making robots more human-like so it is a reasonable to predict that the robot category will show an UVE. The doll and statue categories were included as other examples of non-human objects which approximated human-likeness, and with the intention of comparing the results to those

found by Looser and Wheatley (2010) who found that perceptions of animacy emerged at the near-human point for a set of images which included dolls, statues and robots. They theorised that awareness of whether something is alive or not may contribute to the UVE and so it was predicted that a valley pattern will emerge for these two categories as well. Finally, the animal category was included to consider a particular type of relationship between human and non-human, as it is the only category where the non-human examples are clearly animate and alive but definitely non-human. With the doll, robot and statue categories, the non-human examples all have qualities which are object-like as well as person-like, but this is not the case for the animal examples.

The aim of examining different types of non-human categories was to allow conclusions to be drawn as to whether the UVE occurs due to unique process of making a robot appear more humanlike, or whether it is a more general effect emerging with any transition between non-human and human. However, this study did not find evidence to support the UVE as the strangeness and eeriness rating data did not show a peak in at the 75% human morph stage, and when the human-likeness and strangeness ratings were plotted against each other, none presented a UVE-type curve. However, the results did show differences in the three measures at each level of the morph category, indicating that humanising the non-human agents did affect perceptions of human-likeness, strangeness and eeriness. Each measure also showed an interaction between the degree of human-likeness, as manipulated by the morph stage, and the type of image category that was used, so it can be concluded that the influence of the morph stage on the three ratings was influenced by the type of image that was used in the morph. To consider these differences, the results for each image category will now be discussed.

The robot category represents the archetypal non-human in terms of the UVE, as Mori's prediction was that the dip in emotional response should emerge as these agents

become humanised. There was a gradual increase in the human-likeness ratings of robots as the morphed faces progressed from non-human to human, but eeriness and strangeness peaked at the 50% human point rather than the 75% human point. When the strangeness and human-likeness ratings were plotted against each other, the pattern did not resemble that which would have been expected in the UVE. The robot faces were also rated as quite strange in their un-morphed state, which initially seemed like an unexpected finding, but in a comparison with Mori's original chart in Section 1.1, this may have been as the robots used in this study were all closer to the 'humanoid robot' than the starting point of 'industrial robot' posited there as inhabiting the region of lowest human-likeness.

Dolls were not included in Mori's original graph, but have been anecdotally located within the UV as they can appear eerie and unsettling, and they were also part of the image set used by Looser and Wheatley (2010) in identifying animacy as emergent as the near-human point in a non-human to human morph sequence. Their rating results were different from those predicted by the UVE and also different from the other image categories as they were the only category where eeriness and strangeness decreased as human-likeness increased. This may represent a general property of dolls when transformed into human faces, but it may also be due to the specific images used in this study. The three doll images used as starting points for this study (See Appendix 2d for illustrative images) may have been responsible for this initial positioning of the un-morphed images as highly strange and eerie, as two of them were infant-like, potentially 'cute' dolls that were isolated from any context and their presentation as isolated stimuli against a dark background may have been unsettling due the lack of context rather than any characteristics of the faces themselves.

The statue images were similar to the dolls in that they were initially rated as rather strange, but they showed a small increase in strangeness and eeriness across the morph stages, at least up to the 50% human point, rather than the decrease shown by the dolls. This may be because the statues chosen for this study were manufactured objects designed to present a highly humanlike appearance and in this way they may not have appeared as distinctively non-human, and so could not reasonably span that continuum between non-human and human.

Finally, the morphed images with animal faces as the starting point of the morph were clearly identified as non-human in their un-morphed state, as would be expected, and they were rated as highly strange and eerie at the 50% human point. As there were no suggestions from the literature as to the nature of the relationship between human-likeness and strangeness or eeriness for morphs starting with animal images, this was an interesting novel exploration and the finding that the 50% morphs between the animal and human images were rated as the eeriest and strangest does suggest that there is more to the UVE than can be explained in terms of a continuum between object-like and human-like. The animal and robot categories may be the most interesting of those described here, as they span the greatest distance across the scales of human-likeness and strangeness. Compared to the animals and robots, the dolls and statues both occupied smaller regions in the humanlike-ness/strangeness space. Of the four categories, the animal and robot categories were the two with the highest ratings for the near-human (75%) morph stage. However, it cannot be said that there is an UVE for nearly humanised robot faces, as this study did not find that nearly humanised robots at the 75% point were eerier than the robots at other morph stages. However, it may be that valley in emotional response does exist, but the proposed location is closer to the 50% human rather than the 75% human point, or these results may have been due to experimental effects of

morphing the images in a particular way. These issues will be considered further in Section 4.4, the General Discussion of this Phase. While the results of Study 3 are certainly interesting and notable in their own right, their core purpose was to examine how the stimuli would vary in terms of their human-likeness, strangeness and eeriness to inform Study 4 in looking at the effect of inversion of each of these image types. This study will now be introduced.

4.3: Study 4: The effect of inversion on recognising near-human faces

4.3.1: Introduction to Effect of inversion

Study 4 used an experiment to present images of faces which varied in their level of human-likeness and the type of non-human entity presented. These images were presented in either an upright or inverted orientation and participants' recognition performance was measured. The aim of the study was to allow conclusions to be drawn as to the extent to which each type of face at each level of human-likeness would be subject to a FIE and thus whether there was any support for the hypothesis that NHFs are processed more analytically than human or non-human faces.

Studies of the FIE have considered the impact of inversion on different types of face processing tasks, generally finding that performance at those tasks is decreases when faces are inverted. Two types of tasks were considered for this study, and these were 'pair comparison tasks' and 'same/different' or 'new/old' tasks. The pair comparison tasks were used in the following studies: Yin (1969), Diamond and Carey (1977), Pellicano and Rhodes (2003), Yovel (2005), Searcy and Bartlett (1996) (Experiment 2), Freire et al (2000), Rhodes et al (1993), and Sekuler et al (2004). In these studies, participants learned a target face image and were then asked to identify that target face when it was presented in a pair with a distractor face by comparing the two images and deciding which one was the one

they had seen before. The same/different or new/old tasks were used in Zlotowski and Bartneck (2013), Searcy and Bartlett (1996) (Experiment 1), Freire et al (2000) Experiments 1 and 2, Hills et al (2011) and Schwaninger et al (2009). As with the pair comparison tasks participants were presented with a target face image to learn but rather than picking out the face they had seen from a choice of faces, they were presented with another image and asked to indicate if it was the same as the one they had previously seen. In this type of task, faces are presented singly so no simultaneous comparison is possible.

The relative merits of the task types were considered and it was decided to adopt a pair comparison task for this study, partly as it was felt that Searcy and Bartlett's (1996) study had a particular resonance with the UVE as their stimuli include distorted and disfigured faces so a useful comparison could be made between the present study and their findings. Searcy and Bartlett (1996) used the pair comparison task and a modified version of that task was adopted in this study, with a modification made due to the nature of the faces used in this study, as even though the target and distractor faces were designed to be as similar as possible given the scope of the images used, they were all highly distinctive as shown in the example in Figure 31.



Figure 31: Test image set for an animal to human morph at the 75% human, 25% artificial point.

It was felt that a pair-comparison task could risk a genuine effect being masked by a ceiling effect of very high performance regardless, of the human-likeness or non-human

category. To make the task more difficult and have a better chance of capturing an effect if one exists, a modified version of a same-different task was used where participants were asked to select the training image from three test images. The left to right order of the test images was controlled to ensure that there was no over-representation of the test image in the left, middle or right position. Accuracy and response speed were both measured. In Study 4, with accuracy included as a check on the presence of ceiling effect of identification performance and response speed as the measure for analysis, against which the hypotheses were tested.

Research Question

The theory of a FIE states that faces should be harder to recognise when inverted. Phase 1 suggested that NHFs may be processed analytically rather than holistically, meaning they may not be as affected by inversion as human or non-human faces. These analyses look at the reaction speed results to see if there is evidence for a difference in reaction time when faces are presented inverted rather than upright, to address the following two hypotheses:

H1: (Classic FIE hypothesis) A FIE effect will be found where all faces take longer to recognise when they are presented inverted rather than upright.

H2: (UVE hypothesis) Any FIE effect will be related to the human-likeness of the image, operationalised here as the morph stage of each image. If the NHFs are processed analytically rather than holistically, the speed of identifying the NHFs should be less affected by inversion than the speed to identify the human faces. It is predicted that there will be an interaction between morph stage and inversion, based on the findings of Study 3 which demonstrated that the image category did affect the ratings of strangeness, eeriness and human-likeness at each morph stage. Post-hoc examination of the results data will be used to examine the nature of any difference in FIE for these categories.

4.3.2: Method

Design

The study used a mixed design with one between-subjects factor: morph stage (5 levels: non-human, 25% human, 50% human, 75% human and human) and two within-participants factors (4 levels: animals, dolls, robots and statues) and orientation (2 levels: upright and inverted). The dependent variable was the time to make a correct response, defined as the time it took for the participant to correctly select the previously seen image when it was presented alongside two foil images.

Materials

Stimulus Images

This study used the sixty images of faces as described in *Ratings of morph faces for inversion study above*. They represented the five morph stages created by morphing images of animals, dolls, robots and statues into matched human faces. Their experimental role was to serve as target images for recognition: 120 ‘foil’ images were also created using the same morphing technique to serve as distractors when participants were choosing which was the target image they had previously seen. The foil images were additional examples of each image category and morph stage. The distractor and target images were combined into panes of three images presented side by side. (An example is shown in Figure 32 below.)

Participants were asked to identify the target face by selecting whether it was the leftmost, middle or rightmost face out of the three shown in the testing panel. The foil images were matched for image category and morph stage to the target image. The left to right order of the three faces in the testing image panels was balanced to ensure that the target image fell equally and randomly in each of the three positions.

Inverted version of images

To explore the effect of presenting the images upside-down as well as upright, it was necessary to create inverted versions of the target and test images. The original images were inverted using a batch software process which vertically flipped the images. In this way, the test panels retained the same left to right order as shown in the example below:

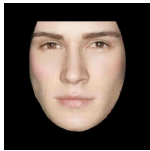

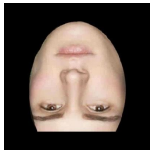
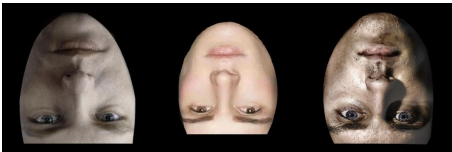
| | Target image | Testing image panel |
|----------|---|--|
| Upright |  |  |
| Inverted |  |  |

Figure 32: Inverted and upright versions of the original and testing images.

Stimulus Presentation Design

Superlab 4.5 stimulus presentation software was used to present the images to participants in a controlled manner and to collect their responses. In this study, separate blocks were created for each of the between-subject conditions: participants were assigned to one of the five morph stage levels at the start of the experiment. They then saw 24 trials involving the same 12 target images in both upright and inverted conditions. The 12 target images comprised 3 examples of each image category but at the same morph stage. Each trial involved first learning the images for 10 seconds (the learning phase), followed by a delay, and then showing the target face and two matched foils (the recognition phase) where participants were asked to select the target face. A training block was included at the start of the experiment to give participants the chance to become familiar with the apparatus and resolve any comfort issues or queries before proceeding to the live trials.

The events were presented in a constrained random order. As the left to right order of the test panel was the same for the upright as for the inverted version, presenting the

upright trial directly after the inverted version (or vice versa) was avoided in case the first experience provided an advantageous cue to the second viewing. While the presentation order was still random overall, it was constrained so that no pairs of upright and inverted images were ever sequenced together.

Apparatus

A 15' MacBook Pro was used to present the experiment. A customised keyboard cover was created specifically for use in this study as a result of the feedback from the pilot study, as it covered the standard keys and indicated which key the participant should press to indicate the choice of the left, middle and right image. Participants were asked to sit at a comfortable distance from the screen, and directed to place their dominant hand close to the keyboard to be able to press the keys quickly. In practice this meant that they were seated no more than 60cm back from the screen, and were able to adjust the rake of the screen to a comfortable angle.

Pilot

Two pilot tests were carried out for this experiment. The first was a quality assurance test to ensure that the pseudo-random ordering of the trials was working correctly where the experiment ran ten times in each condition and the order of the images was manually logged. The logs were analysed to check that the same testing panel did not appear upright and inverted in consecutive trials. The second test asked a volunteer participant to work through the standardised procedure as described below, with the exception that the tester was also given a notepad and asked to write down any comments or problems during the experiment. Two amendments were introduced to improve the clarity of the briefing instructions and to add the keyboard guide to make it easier to see which keys to use to give a response.

Participants

54 participants were recruited from members of Open University staff. Recruitment was carried out using an advertisement on the institution's Intranet noticeboard where volunteers were invited to make contact with the researcher and make an appointment to attend an experiment session. It was desirable that participants were naive with regard to the UVE but essential that they had not taken part in the rating exercise: participants were asked at the expression of interest stage if they had taken part in the researcher's previous projects and were thanked for their interest but screened out if they said that they had. Participant gender and age were recorded at the end of each session. 65% of the participants were female and the mean age was 43 years old.

Procedure

Each participant completed the experiment in a single session in a quiet location. Each had been sent an information sheet (see Appendix 2f) in advance of their session, so on arrival, after being welcomed and asked to take a seat, they were asked if they had read and understood the information that had been sent in advance. Participants were then asked to read and sign an 'agreement to participate' form to read and were asked if they had any queries they had about the research or the process. At this point some participants queried how long the study was likely to take but all were happy to proceed when they were reassured that the duration was likely to be ten to fifteen minutes.

While the participant read and signed the consent form, the researcher launched the Superlab software, assigned the participant to the appropriate morph stage condition by referring to a list of randomly generated numbers to balance participants evenly across the five conditions, and loaded the experiment. From this point on, the experiment ran automatically in Superlab and proceeded at the speed set by the participant and the researcher only intervened if the participant had problems or queries.

The laptop was moved to a convenient space in front of the participant where it was free from glare and shadow, and the researcher demonstrated the keyboard cover and guide, explaining to the participant that during the experiment they would be making choices between several sets of three images to indicate which ones they recognised. Participants were told that there would be two parts to the experiment so they would experience a training phase where they would have the chance to try a practice trial to check they were comfortable with the instructions and the apparatus and a set of twenty-four live trials where their responses would be collected. Participants were told that from now on the experiment would proceed solely on the laptop and that the researcher would not interrupt but would remain nearby to help if there were any problems. They were advised that the laptop would tell them when they had completed the experiment and to let the researcher know so she could complete some final tasks before the session was completed and the participant could leave. A final check was made to ensure that the participant understood the instructions so far and was still comfortable and willing to proceed. They were asked to start by reading and following the instructions on the first screen, which explained that they would be looking at images of faces and then being asked to identify them from a line-up of similar faces, and then to press any key on the keyboard to begin.

The experimenter moved to a nearby chair. In most sessions participants worked through the experiment without any need for intervention or help but a small number of participants had questions after the training example. As this was a reasonably short experiment all participants completed it in a single session and provided data for all the trials. The final screen in the experiment told participants that all the required data had been collected and presented them with a detailed description of the aims of the study and how their participation would be used to contribute to a further understanding of the

UVE When they had read this, participants indicated to the researcher that they were finished. At this stage, several participants had further questions about the specifics of this study or about the UVE in general and these were answered in detail.

At the close of the experiment participants were asked if they were happy to disclose their age and have their gender recorded for demographic analysis purposes. If consent was given these items were recorded against their participant identifier and stored separately from their signed consent form. Finally, participants were asked if they would be interested in receiving a summary of the final results of the project and if so, their preferred email address was collected. Again, this personal information was not associated with their participant identifier. Participants were warmly thanked for their interest and their time and the session was closed. In terms of the duration of the experiment, 30 minutes was allowed for each experiment. Of this, the experiment took around fifteen minutes to run, with the introductory formalities and final question and answer session comprising the rest of the time.

4.3.3: Results

The FIE states that faces should be harder to recognise when inverted. Phase 1 suggested that NHFs may be processed analytically rather than holistically, and so may not be as affected by inversion as human faces. This section analysed participant's reaction speeds to see if there is evidence for a difference in reaction speed when the NHFs from different image categories were presented upright and inverted. Two hypotheses were proposed:

H1: (Classic FIE hypothesis) A FIE effect will be found, where faces take longer to recognise when they are presented inverted rather than upright.

H2: (UVE hypothesis) Any FIE effect will be related to the human-likeness of the image, operationalised here as the morph stage of each image. If the NHFs are processed analytically rather than holistically, the speed of identifying the NHFs should be less affected by inversion than the speed to identify the human faces. It is predicted that there will be an interaction between orientation, morph stage and inversion.

Data preparation

Each participant was allocated to one of the five morph categories and contributed 24 response times from their experience of looking at three images, from four categories, in two orientations. Accuracy was high at 99% across all trials, as only 12 participants made any errors and only one participant made two errors during their session. This indicates a ceiling effect of accurate recognition in this study, so accuracy will not be considered further in these analyses. The 13 errors (from a total of 1296 trials) are summarised in Figure 33 which details the images where the target face was not correctly identified. Statue #3 in its inverted presentation is notable as the image with most errors.




| Image description | Doll #2, Upright, 25% human | Statue #3, Inverted, un-morphed | Animal #3, Upright, un-morphed |
|-------------------|---|---|---|
| Number of errors | 1 | 11 | 1 |
| Image |  |  |  |

Figure 33: Images which were not correctly identified by participants in Study 3.

Corrected means were calculated to replace the response times for the 13 incorrect responses. The corrected response times for the different faces were collapsed to give a single value for each participant, representing their speed of recognising faces in that category in both orientations.

Analysis of response time

Table 11 presents the descriptive statistics for the response times across all morph stages and categories.

| Image category and orientation | Morph stage | | | | | | | | | | All stages mean | All stages SD |
|--------------------------------|-------------|--------|-----------|---------|-----------|---------|-----------|---------|---------|---------|-----------------|---------------|
| | Non-human | | 25% human | | 50% human | | 75% human | | Human | | | |
| | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | | |
| Upright: Animals | 4653.32 | 511.65 | 4913.27 | 664.69 | 5442.30 | 578.82 | 5332.00 | 948.14 | 4726.13 | 808.68 | 5016.68 | 767.80 |
| Upright: Dolls | 4516.27 | 329.33 | 4953.35 | 602.80 | 5343.17 | 441.86 | 5464.08 | 933.94 | 4976.13 | 852.62 | 5054.22 | 737.41 |
| Upright: Robots | 4530.55 | 427.93 | 4609.03 | 436.12 | 5205.93 | 599.63 | 5196.42 | 935.90 | 4900.17 | 946.03 | 4888.02 | 740.66 |
| Upright: Statues | 4400.15 | 306.39 | 4684.91 | 596.98 | 5259.17 | 1013.85 | 5104.97 | 815.13 | 4778.07 | 822.30 | 4843.85 | 779.18 |
| Inverted: Animals | 4692.30 | 440.75 | 4779.67 | 561.09 | 5457.10 | 463.52 | 5510.53 | 1265.33 | 5106.20 | 1232.52 | 5110.20 | 916.90 |
| Inverted: Dolls | 4551.94 | 454.29 | 5230.73 | 925.83 | 5688.90 | 1104.17 | 5685.28 | 833.92 | 4945.50 | 847.98 | 5225.49 | 933.53 |
| Inverted: Robots | 4541.58 | 464.34 | 4867.36 | 1000.07 | 5358.70 | 883.34 | 5823.39 | 2162.82 | 4912.37 | 771.91 | 5112.77 | 1283.54 |
| Inverted: Statues | 4412.55 | 397.32 | 4886.88 | 656.38 | 5156.23 | 436.93 | 5179.39 | 820.26 | 4844.83 | 829.63 | 4897.35 | 692.82 |

Table 11: Mean response times and standard deviations for each morph stage, upright and inverted.

The response times were analysed using a three-way mixed ANOVA to explore image category and morph stage at both orientations. Mauchley's test indicated that the assumption of sphericity had been violated for the main effects of category $\chi^2(5)=36.38$, $p < .0001$, $\epsilon=.66$ and the interaction between category and orientation: $\chi^2(5)=21.61$, $p < .0001$, $\epsilon=.79$. As a consequence, Greenhouse-Geisser corrections to the degrees of freedom were used in interpreting those significance tests. The between-subjects main effect of morph stage was found to be significant: $F(4,49)=3.04$, $p = .026$, partial $\eta^2=.199$. The two within-participants main effects were also found to be significant: image category ($F(1.98,96.76)=5.84$, $p < .05$, partial $\eta^2=.106$) and orientation ($F(1,49)=10.77$, $p < .05$, partial $\eta^2=.180$). None of the interactions were significant: morph stage x orientation, ($F(4,49)=1.124$, $p = .356$, partial $\eta^2=.084$) category x orientation ($F(2.39,116.936)=.586$, $p = .588$, partial $\eta^2=.012$) and category x orientation x morph stage ($F(12,147)=.782$, $p = .668$, partial $\eta^2=.060$).

The main effects of orientation, image category and morph stage will now be considered to evaluate evidence in support of H1 and H2.

Orientation

Upright faces were identified significantly faster than inverted faces ($M=4949.47\text{ms}$ for upright faces, compared to 5081.57ms for inverted faces.) This supports H1, that faces would be recognised faster when upright than inverted.

Morph Stage and Image Category

H2 predicted an interaction between orientation, image category and morph stage, where the NHFs would be faster to recognise when inverted, compared to the human images. This interaction was not found. Significant differences were found between the image categories and the morph stages. Post-hoc tests of four simple pairwise comparisons using Bonferroni's correction were used to examine which of the morph stages differed significantly from each other. Only one of the comparisons was significant, with a significant mean difference in response speed found between the Artificial and the 75% human morph stages, with a 874.67ms slower mean response speed for the 75% human faces, $p < 0.05$. Three simple pairwise comparisons using Bonferroni's correction were used to look at the image categories. Two of those comparisons were significant, with the statues identified significantly faster than the animal or doll faces (statues were recognised 191ms faster than the animal faces, $p < 0.01$, and 265ms faster than the doll faces).

The slower response speed for the NHFs compared to the human faces was unexpected. Figure 34 and Figure 35 below show the mean response speed for each image category and morph stage.

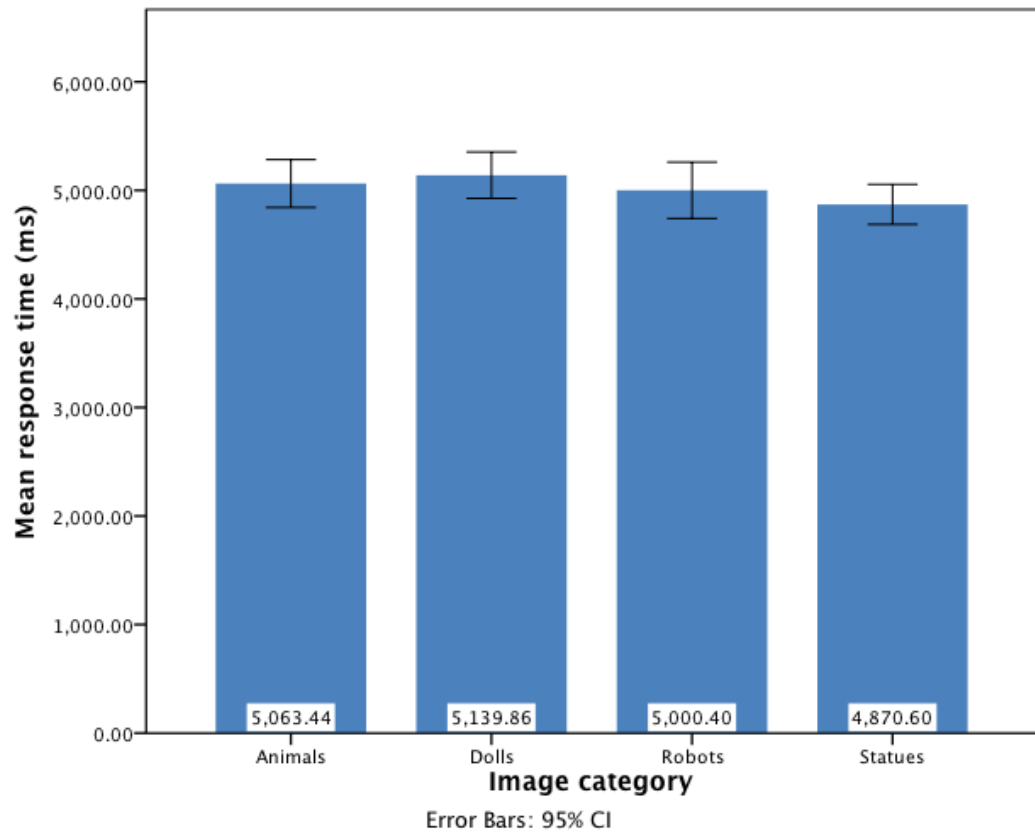


Figure 34: Mean response time for each image category.

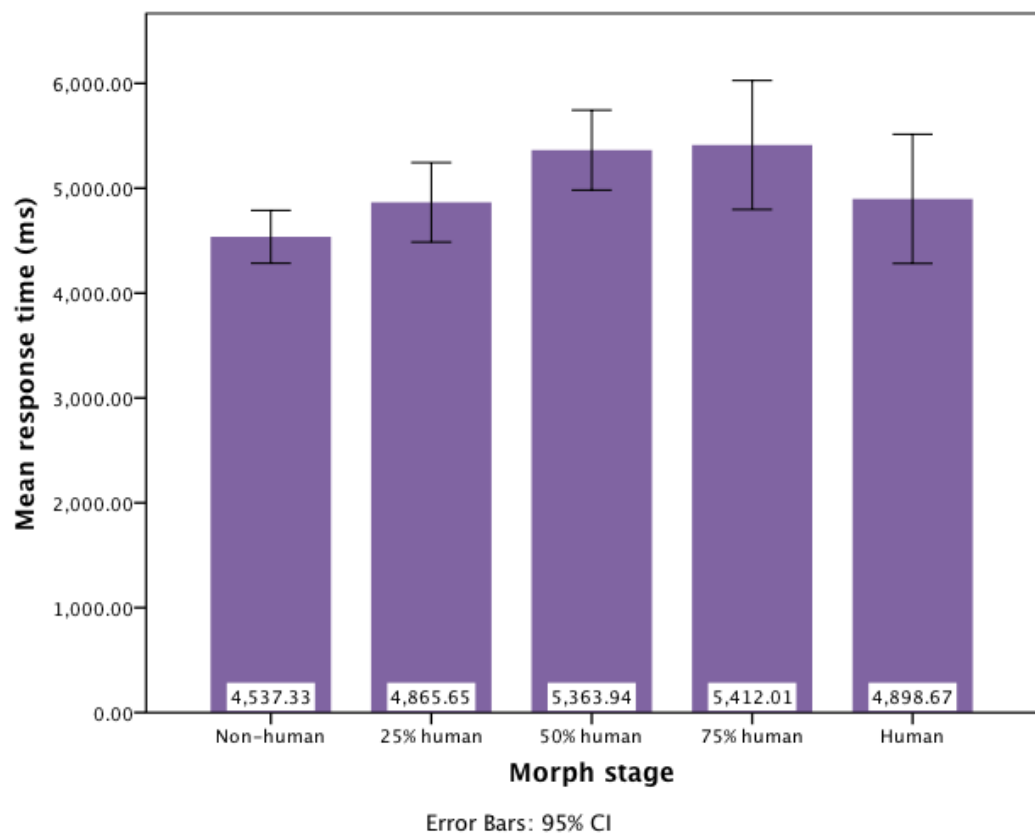


Figure 35: Mean response time for each morph stage.

4.3.4: Discussion

The research question in Study 4 was to investigate whether a FIE would be found for a collection of faces which had been systematically varied in two ways. They were created by morphing artificial faces into human faces so the first variation was the level of human-likeness, measured by the point in the morph sequence that each image had been taken from. The second variation was the nature of the non-human starting point of the morph, to explore whether there would be a difference in the responses to morphs created from animals, dolls, robots and statues. Classically, Mori's (1970) UVE describes an emotional response that occurs as robots are made more human-like. However, as there are several other types of non-human entity which can also be humanised, three other categories were included to compare whether any FIE would be affected by the type of non-human starting point. This is part of a general enquiry as to the nature of the UV and to explore whether the UVE is an artefact of a particular type of artificial to human transformation or if it is a more general phenomenon.

This study was grounded in face perception research into how images of faces are perceived and draws upon the long-standing finding that face perception performance across a variety of tasks is degraded when faces are inverted. This disruption in perception is generally compared to how non-face objects are perceived, as they are generally seen in terms of their component parts rather than as a holistic entity, so inversion does not have such a disruptive effect on perceptual ability. This theory was applied to NHFs to test a hypothesis that the perception of these faces may be different to that described for normal human faces as studies by Brenton et al (2005), Geller (2008) and Seyama and Nagayama (2007 and 2009) all observed that a characteristic of faces thought to be uncanny was that their eyes were unusual in some fashion, either exaggeratedly large or

unusual in their appearance or positioning. If analytic processing is already being engaged for these NHFs, inversion should not cause a degradation in performance so a smaller effect of inversion should be found.

Participant's reaction time and accuracy of response were measured in this study. Accuracy rates of 99% were found so no further analyses of accuracy were conducted. The analyses in this section looked at the differences in response time across the different morph stages and categories at each orientation, to test two hypotheses. The results were supportive of H1's prediction of a classic FIE, as across all categories and morph stages, response time was significantly slower for the inverted faces compared to the faces presented upright. However, there was no support for H2 that the NHFs would be processed more quickly than the human faces. The opposite effect was found with the NHFs processed significantly more *slowly* than the artificial faces. The 75% morph faces took significantly longer to identify compared to the artificial faces, but no significant difference was found between the human and artificial faces, or between any of the other morph levels. The image category also had a significant effect on the response time as dolls were found to have the slowest response time and statues the fastest. Considering the individual images that were selected for inclusion in this study, it may be the case that the three doll images subjectively appeared more animated and 'lifelike' than the statues, which may suggest that the statue images are being perceived in terms of their object-like nature and this is not occurring for the doll faces as they are manufactured objects designed to look both humanlike and life-like. The significant differences in response times to the artificial and NHFs does provide evidence that the faces in these two morph stages are being processed differently but the nature of that difference was not as predicted as they did not show a decreased inversion effect, but instead the response times were slower, rather than faster.

It may be that some aspects of the research design inhibited the ability to demonstrate the UVE, and a critical reflection on these does suggest ways in which this study could have been improved. The nature of the stimuli used in the trials meant that it was very difficult to produce sets of images that were relatively homogenous but still allowed the morphing approach to subtly vary the human-likeness of each image, with the result that the test images and foils did naturally look quite different from each other. It is useful to reflect on the extent to which this meant that participants were able to apply pattern matching and the identification of salient characteristics within the target image to identify them from the test foils. This is acknowledged as a limitation of the type of design used for this study, and one way in which it could potentially be addressed would be to make a transition away from using photographs of humans and non-humans as the base images and instead, working from the start with a more homogenous set of CGI stimuli. At the time when this study was developed, the image quality of these was too poor for participants to believe that any might actually be realistic humans or realistic animals, dolls, robots or statues, but progress in this area means that it may now be a realistic proposition for follow-up work, and it would help to smooth out that marked difference between the appearances of the different images.

4.4: General Discussion of Phase 2

This section will review the findings from two studies included in Phase 2 and discuss what can be concluded from their findings. This phase was designed to explore the premise that one aspect contributing to the UVE may be that the NHFs which trigger an unsettling or eerie response may also display an unusual property in that they are processed analytically rather than holistically. This premise was grounded in previous research into the UVE which suggested that unusual or distorted eyes may contribute to

an impression of uncanniness, and then developed with reference to psychological research into face processing which has explored evidence of analytic versus holistic processes for different types of face. It was designed as no studies have been carried out which systematically varied the human-likeness of a face and measured whether speed of discrimination would vary as the face became closer to human.

This phase used a morphing technique to generate a set of faces which represented five points on a continuum between non-human and human, created from four different categories of non-human face. If the NHFs are processed analytically, it was expected that the faces at the 75% point would be less affected by inversion than the natural human faces or the artificial faces. Participant ratings of their human-likeness, strangeness and eeriness were gathered to ensure consistency with studies presented in Phase 1 and also to confirm whether these did represent a reasonable continuum of human-likeness. These ratings were also analysed by image category to explore whether the UVE emerges as a general effect when any non-human face is made more human-like, or whether it is a specific property of robots increasing in human-likeness as had been originally proposed by Mori (1970). Analyses of these ratings found that human-likeness did indeed increase as the morph stages progressed from non-human to human, but that there was not an entirely consistent increase between each stage as there was a larger difference between the 50% and 75% stages than between any of the other stages. The ratings of strangeness and eeriness did not provide support for the UVE hypothesis as it had been predicted that the faces at the 75% stage would be rated as the strangest and the eeriest but it was found that the mid-point face, representing 50% human and 50% non-human, was rated as the strangest and eeriest overall. The finding that the 50% point was the peak of eeriness and strangeness for several of these images was unexpected, particularly for the robot images. The images did vary systematically in human-likeness, but as observed

above, participant ratings of human-likeness showed a large increase between the 50% and 75% human point. It may be that the real peak for eeriness and strangeness would actually have emerged at a point between 50% and 75% if the image set had included morphs from those stages. A suggestion for future research would be to replicate the study with the same initial and end-point images but taking more than just the five points along the continuum to evaluate the ratings for a larger number of points along the morph. This would allow a more detailed evaluation of the human-likeness, strangeness and eeriness than was possible with only five points.

These images were then used to test whether the degree of human-likeness or the image category would affect how quickly participants responded to a recognition task, when the faces were presented upright or inverted. This was to test the hypothesis, grounded in the FIE literature, that participants would respond more slowly when some of the faces were inverted. Previous studies (e.g. Farah and Tanaka, 1995; Yin, 1969) suggested that performance should be significantly degraded when trying to recognise the natural human faces upside-down as this presentation interferes with the configural processing which is generally engaged in recognising a face. Yin's initial description of this FIE found that this did not occur with non-face objects of comparative complexity, and theorised that this may be because participants pay selective attention to individual component features of those objects and these serve as recognition cues as they can be identified when the object is inverted. The key hypothesis for this study was derived from the findings of Phase 1, which suggested that participants may perceive NHFs analytically rather than holistically, and that participants may treat the NHFs in the same way as non-face objects, encoding them analytically because of the distinctive features that also make them appear unsettling or eerier. If that was the case, then these faces would not suffer from the same degradation in performance as the natural human faces as they are

already being processed in an analytic fashion which is not disrupted by inversion. It was predicted that the NHFs would show a smaller inversion effect compared to the human faces, but this effect was not found as there was no interaction between orientation and morph stage. Overall, a FIE was found as participants were indeed slower to respond to the inverted faces compared to the upright faces but the predicted advantage for the NHFs was not found. It had been expected that the NHFs would have the fastest response speeds when presented inverted. Participants were actually significantly slower to respond to those faces. One possible explanation for this may be that the original prediction did not take into account the concept that inversion has been found to disrupt the perception of some types of manipulated faces differently to others. Searcy and Bartlett (1996) investigated two types of manipulated faces to look for evidence of whether holistic versus analytic modes were being used to encode the information present in the face. Their design created grotesque faces, some of which were manipulated by 'disfiguring' the individual features and some by changing the location of the features in the face. In this way, it was possible to clearly distinguish between those spatial changes that affect the holistic impression of the face, and component changes which would affect the analytic encoding. As with the present study, they measured the extent to which performance was degraded when these differently manipulated faces were inverted and found that inversion did not impair the performance equally for the two types, as inversion had a more marked effect for those faces which had been spatially manipulated but very little effect on those that had distorted components. The morphs created in the present study were designed by matching the positions of facial features (eyes, nose, mouth, and hairline where possible) and then digitally altering the rest of the face around those fixed points. The technique was chosen as a method of systematically varying human-likeness but it also introduced an additional variation, that of gradually

distorting the components of the face while keeping the spacial relations between them constant.

Some limitations of the studies in Phase 2 have also been noted above as the ability to draw detailed conclusions about the differences between the levels of human-likeness is limited by only having the five morph stages, and concerns have been noted on the variance in rated human-likeness between the different morph stages. These suggest areas for follow-up research to address those issues. Future research should also consider whether morphing is still the best technique for systematically varying the human-likeness of an agent. The techniques used in the present study did allow a systematic transformation through from artificial to human, but it does have the disadvantage that morphed images tend to be very clearly identifiable as the product of a blend between different images. In addition, it is useful to consider the issue of anchoring in the context of these studies. In both studies, the non-human and human images were included as a way to provide empirical end-points on a continuum of images of human-likeness which started with something that was very clearly not a human being, and progressed to images which were clearly and definitely live human beings. This intention was grounded in the design principles described in Table 1, where being able to reliably place entities on an empirical continuum was seen as a key component of studies which should be able to draw conclusions about the UVE. However, it could well be argued that simply including these images in the set to be rated or tested does not fully capture human abilities to respond to images that vary in human-likeness, as each image was presented in isolation and without any possibility of comparison to other images in the set. It would be useful and interesting to compare whether either of these studies would produce the same findings if replicated using a comparative presentation where the anchor points were

visible and constant, and used as deliberate reference points for the different levels of human-likeness.

This phase has provided some evidence suggesting that there are differences in how faces at the 75% morph stage are processed compared to non-human faces, but the anticipated difference for NHFs compared to human faces were not found. The notable finding in terms of the UVE was the identification of the 50% artificial and 50% human faces as the strangest and most eerie which serves to challenge the original prediction that almost but not quite human faces would be the 'uncanny' candidates. The studies in this phase applied theories and methods from the psychology of face perception to help understand the UVE and ground it in the context of established research. It has addressed one of the initial strands of enquiry for this thesis by looking at the cognitive aspects of the UV, a strand that was initially developed from observations in the description task of Phase 1. Phase 3 also developed the findings of Phase 1, but in a different direction, by looking in more detail at the role of emotional expression in the perception of NHFs as uncanny.

Chapter 5: Phase 3: Study 5: Facial Expressions of Emotion

5.1: Introduction to Phase 3

Mismatched expressions and the Uncanny Valley Effect

This final research phase considered whether our reactions to NHFs can be explained by subtle mismatches between expectations that are set up by the overall appearance of an agent and the detail of the emotional expressions it displays. In particular, it tested whether particular types of mismatch contribute to triggering an UVE. Integral to the uncanny valley concept is the premise that during the “artificial to human” trajectory, an agent moves from merely looking lifelike to seeming convincingly alive and transforms from being predominantly object-like to eventually appearing human. With increasing human-likeness, it would be reasonable to assume that our ability to understand, predict and empathise with the agent would also increase but the UV dip appears to be present in levels of affinity as well as acceptability, which results in the NHAs being rejected. This phase proposes that this rejection may occur because the face is able to convey an overall impression of human-likeness and realism that cannot be matched by the realism of the expression that the agent portrays. This makes it very difficult for any potential social connection to be achieved between the viewer and the agent and causes rejection of the NHA. To use an example presented in an earlier part of this thesis, androids such as Repliee Q1 may be unsettling because their movements, gestures and expressions are highly lifelike but their eyes lack the necessary build detail to be able to look completely human. Moreover, the contrast between something appearing highly lifelike in some ways and un-lifelike in others may be deeply unsettling. Mori’s original (1970) paper was recently re-translated (2012) and this new translation is also relevant to consider here. It updated the original ‘familiarity’ term used to label the emotional response axis of the

chart with a new term of 'shinwa-kan' as a more precise way of conveying Mori's intended concept. There is no direct translation into English for this term and while 'familiarity' is certainly a component of the emotion, shinwa-kan is a broader positive feeling which goes beyond the identification of something as familiar to also include notions of affinity, empathy, comfort and likableness. This new term makes more sense of the negative-familiarity region in the uncanny valley chart to explain why Mori thought that certain examples would be eerie. There is no intrinsic link between something being unfamiliar and eeriness as in most cases, novelty is not unsettling in its own right. However, it is easier to see how something which is uncomfortable or unlikeable (negative shinwa-kan) could be unsettling.

With this broader notion of affinity or empathy as the response that experiences a plunge when a NHA is encountered, questions of how emotional expressions are conveyed and perceived in NHAs become particularly relevant. The importance of accurate and realistic emotional expressions in interpersonal relations has been discussed in the Section 1.5 above, but with specific reference to NHAs, this has also been demonstrated from published research on human-robot interaction. Breazeal (2003) found that spontaneous emphatic reactions occurred in participants when working with a robot that could pose different emotional expressions even when the overall appearance of the robot was minimally human. Leite et al's (2013) work on the influence of empathy in human-robot relations found that a robot programmed to display convincing and appropriate facial expressions during an interaction was rated as more acceptable and friendlier than when it made the same responses without the facial expressions. They suggest that establishing empathy through appropriate and believable facial expressions is a requirement for making a meaningful relationship with a NHA. It follows that an inability to do this convincingly due to a mismatch between a general impression and a specific

expression could lead to a rejection of the agent and may explain some of the unease that characterises the uncanny valley effect.

Rationale for expression mismatch study design

The approach used to investigate this question has been developed from findings in Phase 1, as well as from evidence from published studies. In the emotional response task carried out as part of Phase 1, it was found that there were more references to eyes for the barely-humanoid faces than to any other feature, suggesting that this prominent feature was responsible for how participants felt about its appearance. In addition, several of the participants' comments for the barely-humanoid face expressed disquiet at the blank eyes. For example:

'VERY creepy (made me jump when picture loaded) doll with no eyes, just big black holes, and ridiculously small mouth and chin if you can get past the freaky eyes.'

'Creepy, hollow eyes, horror film puppet'

'Go to the figure that looks really freaky!! [...] She has big eyes which are blackened and look like blackened closed doll eyes.'

While that particular face did indeed have prominent and unusual eyes, it was felt that these observations warranted further exploration. The unsettling sensation produced by almost-human faces may occur because the overall impression of the face sets up an expectation that a particular emotion is being expressed which is then violated when that expression is not carried through to the eyes, either because they have been deliberately left 'lifeless' or because the ability to reproduce and animate the expressiveness of human eyes lags behind the ability to produce photorealistic skin texture or bone structure. This mismatch between the expected expressiveness and the expressiveness that can be portrayed may contribute to the sense of there being 'something wrong' about the almost-human entity.

Published research also pointed to the role of unusual or blank eyes in contributing to the UVE, particularly where they are presented in a face where the skin has a realistic texture causing a mismatch between the realism of the eyes and the surrounding face. The background to research into cue mismatches and their role in the UVE has already been described in the Section 1.3 above, but this introduction will recap some key studies before introducing the research questions that were addressed in this phase.

A precedent for investigating a mismatch in aspects of how an almost-human entity is presented can be seen in a study by Creed and Beale (2008) where participants viewed a CG figure whose facial expression and voice inflection could be congruent or incongruent. Participants rated the figure on pairs of positive and negative characteristics and their ratings of the positive characteristics were amplified where those characteristics were congruent. They concluded that being presented with mismatched stimuli did have an impact on participant's responses to the figure and suggested that this may be due to a type of cognitive dissonance. Seyama and Nagayam's (2007) systematic manipulation of artificial and photorealistic faces found that the eeriest faces were those where oversized eyes were paired with realistic skin texture. Similarly, MacDorman et al (2009) found that the eeriest faces were those where the skin texture was photorealistic and the eyes were magnified to 50% their original size. In both cases there was a mismatch between the realism of the overall face (i.e. skin texture) and the expected proportions of the eyes. Tinwell et al (2010) presented participants with videos of faces posed to display particular emotions or a lack of emotion, and compared ratings of human-likeness, strangeness and eeriness. Their key finding was that a lack of emotional expression in the upper part of the face elicited feelings of uncanniness, particularly when a negative emotion was being portrayed in the rest of the face.

This final phase built on the results of those published studies but developed the mismatch idea to look specifically at reactions to mismatched emotional expressions. This develops Creed et al's research into mismatched expressions, as well as Tinwell et al's approach, as the latter studies paired a lack of emotion in the upper part of the face with a different expression in the rest of the face, whereas the present study explored the ratings that would be given for, for example, a face posed to display a happy expression overall but with the eye region from the happy face replaced with the eye region from a face posed to display a frightened expression. These expression combinations are not possible for anyone to produce naturally and so these images do belong to the category of artificially created 'near-human' faces. The original faces from which the expressions were created are also included in the stimulus set so it is possible to compare responses to these NHFs to their human originals. Examples illustrating three of these expression blends are shown in Figure 36 below.

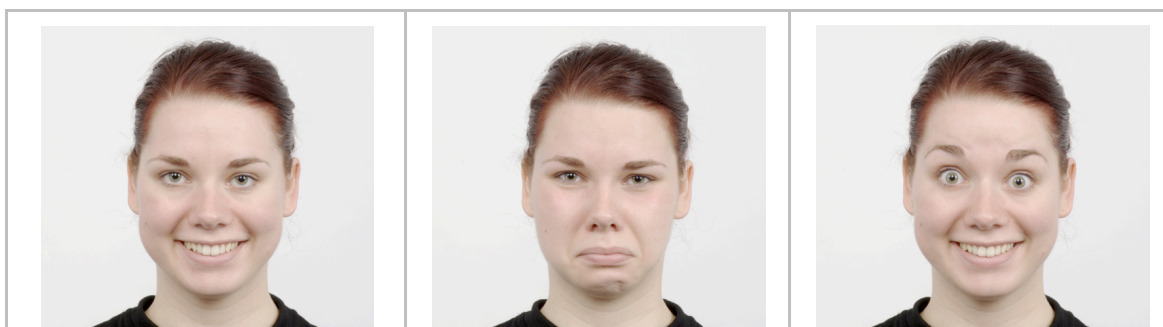


Figure 36: Three examples of face expression blends, from left to right: *happy face, neutral eyes*; *sad face, happy eyes*; *happy face, fearful eyes*.

The premise that some of these combinations may give rise to a response pattern in keeping with the UVE will be explored through topics relating to the perception of emotional expressions and their relationship to the UVE. These have been grouped into four explorations, firstly looking at the core question of the eeriness of the different blends and then going on to explore the differences in how each blend was perceived. This phase used an online experiment to explore whether mismatched expressions are

eerie and whether different types of blended faces resulted in different ratings of eeriness. Strength of participants' emotional responses to the different blends was also measured as well as whether they were able to accurately categorise and recognise the emotions present in each blend. Image orientation was also varied in one of the tasks to explore whether the previous evidence of NHFs being processed analytically rather than holistically could be supported.

The effect of face blend on eeriness: are mismatched expressions eerie?

This exploration looked at the effect of face blend on eeriness. This is the first study to explicitly measure this effect of mismatch on eeriness, so three scenarios for the pattern that the ratings of eeriness may take have been predicted from previous studies whose results touch on similar aspects of face perception.

Firstly, Tinwell et al's (2013) research found that a lack of expression in the upper part of the face was linked to participants feeling a heightened sense of eeriness when looking at particular faces. Participants also attributed psychopathic traits to the entities displaying this expression. If the present study was to support Tinwell et al's finding, the face expression blends where neutral eyes were presented in disgusted, fearful or sad faces would be rated as the most eerie. The angry faces with neutral eyes have also been considered in this hypothesis, but Tinwell et al's (2011b) findings suggested that a lack of expression in the upper part of an angry face did not reliably predict eeriness or negative evaluations, perhaps because cold, angry expressions are not particularly unusual.

Secondly, Ekman et al's (2005) study of 'leaked expressions' found that an attempt to cover up a negative emotion using a forced or fake smile rendered people untrustworthy due to the discrepancy between the smile and the negative emotional cues that were also present in the expression. Ekman et al's work was not an exploration of the UVE but his findings can be applied to the concept because one plausible explanation for the UVE may

be that some NHFs are negatively evaluated as they lack congruent expressions and this may make them appear untrustworthy. In this scenario, it would follow that the happy faces blended with angry, disgusted, fearful or sad eyes would be rated at the most eerie.

Finally, Seyama and Nagayama's (2009) study found that enlarged eyes were a factor in triggering a sense of the uncanny in artificial faces. In terms of the face images used for this study, the blended faces including a fearful expression would display larger eyes because these are widened in fright whereas the angry, disgusted and sad expressions would not be characterised by a widening of the eyes. Therefore, the blends with fearful eyes will be rated as the most eerie and of those, the neutral face with the fearful expression will be rated as eeriest.

Based on these studies, it was hypothesised that there would be significant difference in eeriness, with the mismatched faces rated as eerier than the matched faces (H1). In terms of the nature of that difference, three scenarios were considered. Firstly, after Tinwell et al (2013), it was predicted that the eeriest faces would be those with neutral eyes and disgusted, fearful or sad faces. Secondly, after Ekman et al (2005) it was predicted that the eeriest faces would be those with a happy face paired with any negative eye regions. Finally, after Seyama and Nagayama (2009) it was predicted that the neutral face, fearful eye blends would be rated as eeriest.

A second hypothesis was considered to explore the effect of inversion on ratings of eeriness. Study 4, reported above, found that NHFs were subject to an inversion effect and that the 75% human faces took significantly longer to identify than the non-human faces, suggesting that they may be processed differently. H2 predicted that orientation would have a significant effect on participant's ratings of eeriness, and that inverted faces would not be found to be as eerie as the upright faces. This prediction is grounded in the literature presented in Section 1.5, as Fallshore and Bartholow (2003) found that it was

harder to recognise most negative emotions from inverted faces, and Prkachin (2003) found that negative emotions were more easily confused when viewed in inverted faces.

Face blends and emotional responses: which faces will evoke the strongest ratings of anger, disgust, fear, sadness, happiness and surprise?

As well as being rated for eeriness, the faces were each rated for how strongly participants felt six emotions when looking at each blend. These responses were collected to understand more about the emotional composition of the UVE, as it is generally described as a sense of eeriness or unease but the precise nature of it as an emotion is not yet clearly understood. Ho et al's (2008) findings suggested that both fear and disgust were key components of the UVE. In addition, sadness and happiness were considered under (H3). It was predicted that the mismatched faces, which were predicted to be eerier than the mismatched faces, would evoke the strongest ratings of fear, sadness and disgust and the lowest ratings of happiness.

In addition, as each of the blended faces was produced as a composite of two emotions, the emotion experienced by participants could be one of those emotions or the blended face may elicit emotions not present in either of the original expressions. Identifying which emotions are felt most strongly when looking at the faces, and whether these are congruent or incongruent with the emotion shown in the eyes or surrounding face will allow conclusions to be drawn about the effectiveness of the face as a convincing portrayal of those emotions. H4, based on Bassili's (1979) work, predicts that happiness and disgust will be experienced more strongly when these expressions are present in the eyes of the blend, anger and fear in the surrounding face, but that there will be no such distinct pattern for surprise or sadness. This was explored by inspecting the mean ratings for each of the matched and mismatched blends.

A final question in this section looked at how subjective ratings of eeriness for each of the faces correlated with measurements of feeling angry, disgusted, frightened, sad, happy or surprised when looking at each image. H5 proposed that ratings of eeriness should correlate strongly and positively with participant experiences of fear and disgust but not with sadness or anger. Fear and disgust have been previously anecdotally identified as emotions implicit in the uncanny valley effect, and this is also supported by the work of Ho et al (2008) who found that these were the terms most closely correlated with 'eerie' and 'creepy' in their inventory of the characteristics of NHAs.

Emotion Identification: are there differences in how each face is perceived?

Having explored the ratings of eeriness for each face and looked at correlations between eeriness and experienced emotions, the third area explored the accuracy with which emotions were identified from each of the blended faces to see whether participants were able to identify happiness, neutrality, anger, disgust, fear and sadness when present in the blends. It was generally hypothesised that matched faces would be recognised more accurately than mismatched faces, and four hypotheses were considered for the nature of that difference.

Firstly, Calvo and Lundqvist's (2008) findings suggested that participants may find disgust and anger hard to distinguish and so disgust may be identified when a blended face contains anger and vice versa (H6).

Secondly, the work of Calder et al (2000) suggested a pattern of recognition errors as a function of whether the emotion is displayed in the eye-region of the blend or in the surrounding face. More errors in identifying disgust and happiness will occur in the blends displaying these in the eye region, and more errors in identifying anger, fear and sadness will occur when these are presented as the face expression with mismatched eyes (H7).

A third pattern is grounded in earlier research by Hanawalt (1944) which suggests a broader division where the positive emotions are identified from the top part of the face and negative emotions from the lower part. This would mean that happiness would be identified in the happy-eye blends, and the negative emotions (anger, disgust, fear and sadness) from the blends where those are the expression in the surrounding face (H8). Happy faces with sad eyes present a challenge to Hanawalt's evidence and will be considered with interest.

Another question concerns whether the identified emotions will be the ones displayed in the eyes/top of the face or the surrounding face/bottom of the face? It is possible to create 'chimeric' faces by taking the top part of a picture of a face and aligning it with the bottom part of another image. Calder (1996) used this technique to create chimeric faces where two different expressions are presented in one face. While the blends in the present study have been created by isolating and transferring only the eye region from each face leaving the rest of the top of the face intact, chimeric faces do have obvious similarities with these blends and the findings from that literature have been used to inform hypotheses about the way in which emotions will be perceived from the blends. A theme of interest for this study is whether participants would identify particular emotions in the top and bottom halves of the face. H9 predicted that happiness and disgust will be identified in the faces where these are present in the eyes of the blend, and anger and fear identified where these are present in the surrounding face. There will be no distinct pattern of localisation for identification of surprise or sadness. This has been predicted from Bassili's (1979) work (H9).

Emotion Classification: Will the blended faces be classified in terms of the emotion displayed in the upper part of the face, the lower part of the face, or will they be judged neutral or too difficult to classify?

A classification task has been included in this study to see if there are any patterns in participant's ability to classify the blended faces in terms of the emotions they displayed. These results were explored descriptively and three patterns were investigated. Firstly, it was proposed that the eerier blends would be hardest to categorise. Secondly, as noted above, Hanawalt (1944) suggests that positive emotions are generally perceived from the bottom of the face, negative from the top, so the negative eyes, happy face blends would be classified as displaying positive emotions and the happy eyes, negative face blends will be categorised as displaying negative emotions. Finally, the classifications were also considered in light of the first scenario explored in H1 above to see whether the neutral-eyed blends would be particularly hard or easy to classify.

5.2: Method

Design

An independent groups design was used and nine dependent variables were measured, eight in the first task and one in the second. The first task involved two IVs, orientation and face type. Participants were presented with images of faces and asked to rate how strongly they felt six emotions. They were then asked how eerie they found each of the faces, and then to identify the image that they thought the face was displaying. Half of the faces were presented upright and half presented inverted. The second task involved the IV of face type only, and did not test orientation as all images were presented upright. In this task, participants were asked to classify a subset of the face images in terms of whether they presented positive, negative or neutral expressions.

The stimuli used in these tasks were 24 photographs of faces displaying emotional expressions. 18 were blended faces, or Facial Expression Blends (FEBs) defined as a type of chimeric face created from images of posed expressions where the eye region of one face is blended into a base face displaying a different expression. 6 of the images were the original un-blended emotional expressions, 18 were different types of blended images. All combinations are a blend of either a positive, negative or neutral expression with another positive, negative or neutral expression. There are four variants of negative expression: anger, disgust, fear and sadness. These combinations created 24 different levels of FEBs. Figure 37 presents the 24 levels. The faces in the top-left, centre and four bottom-right cells are unmodified images of the original posed expressions. In all the other cells, blended faces have been created by taking the eye region from one expression and blending it into the base face of another.














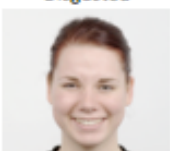
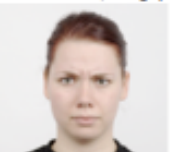
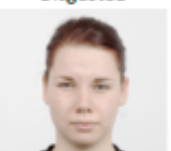
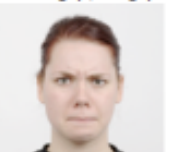
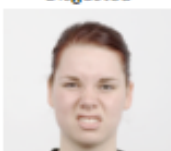






| | | Eyes – eye expression constant across row | | | | | |
|---|----------|---|---|---|--|---|---|
| Face – face expression constant down column | Positive | Positive | | Neutral | | Negative | |
| | | 1 Happy/Happy  | | 2 Neutral / Happy  | | 3A Angry / Happy  | 3D Disgusted / Happy  |
| | | | | | | 3F Fearful / Happy  | 3S Sad / Happy  |
| | Neutral | 4 Happy / Neutral  | | 5 Neutral / Neutral  | | 6A Angry/Neutral  | 6D Disgusted / Neutral  |
| | | | | | | 6F Fearful / Neutral  | 6S Sad / Neutral  |
| | Negative | 7A Happy / Angry  | 7D Happy / Disgusted  | 8A Neutral / Angry  | 8D Neutral / Disgusted  | 9A Angry / Angry  | 9D Disgusted / Disgusted  |
| | | 7F Happy / Fearful  | 7S Happy / Sad  | 8F Neutral / Fearful  | 8S Neutral / Sad  | 9F Fearful / Fearful  | 9S Sad / Sad  |

Figure 37: Examples of all 24 levels of FEBs showing how the expression images were combined.

Table 12 below summarises the design used for the separate parts of the study, presented in the order in which participants experienced each task:

| Task | IVs | DVs | Measure | Survey Question |
|------|---------------------------------------|-----------------------------|--|---|
| One | Face expression blend and orientation | Emotion strength: Afraid | Response on a nine-point scale. The scale was labelled at the left, mid and right-most ends with: <i>I don't feel this at all, I feel this moderately strongly and I feel this extremely strongly.</i> | "When looking at the person in the picture, how strongly do you feel each of the emotions below?" |
| | | Emotion strength: Angry | | |
| | | Emotion strength: Disgusted | | |
| | | Emotion strength: Happy | | |
| | | Emotion strength: Sad | Response on a 9 point scale labelled at the endpoints as <i>not at all eerie</i> through to <i>extremely eerie</i> . | "How eerie do you find the person in the picture?" |
| | | Emotion strength: Surprised | | |
| | | Eeriness rating | | |
| | | Emotion identification | One of the following: <i>Anger, Disgust, Fear, Happiness, Sadness, Surprise or No Emotion</i> | "Which of the following best describes the emotion shown by the person in the picture?" |
| | | Emotion categorisation | One of the following: <i>Positive emotion, neutral emotion, negative emotion or I can't decide.</i> | "Which type of emotion is being shown by the person in the picture?" |

Table 12: Structure of tasks, IVs, DVs, measures and survey questions.

To ensure a highly standardised set of stimulus materials, the images used to create the blends were carefully selected to ensure they were matched for quality and isolation of the face image from any distracting background, lack of facial accessories including glasses or facial hair. Male and female models were included and models of different ages were selected where possible. To avoid the issue of comparative ratings, the image sets seen by each participant were constructed to only present one blend from each model: participants who saw the 'model one' face where the happy expression had been modified by the application of fearful eyes would never see 'model one' displaying any other type of blend. Participants were randomly allocated to one of eight groups. The groups were based on a collection of FEBs structured to ensure that each participant saw every type of blend once but that the type was only seen once in each task and each model was seen only once in the course of the study. Images were presented in a random order within each group.

Pilot testing

Pilot tests were used in the development of the survey web pages and many of the changes involved incremental improvements to the layout of the survey pages and wording of questions but five substantial changes were also made and will be described. The version sent to pilot testers was a functionally complete version of the survey but did not include all potential FEBs. The following changes were made:

- **Input options, Sliders vs Radio Buttons:** An early version of the survey used 'sliders' to collect rating data. Participants operate these by dragging a marker to the correct place on the scale. While they are visually appealing and can add interest to the participant experience, it was concluded that they do have several disadvantages in a study of this nature. They can be hard to operate on mobile devices, meaning that inaccurate response data may be collected. Radio buttons are a less interactive but more accurate input method, particularly where the scale is reasonably short as in the present study. It was decided to adopt radio buttons rather than sliders for the final version of the study.
- **Image Size and Survey Performance:** it was found that the survey was taking some time to load for some of the testers. Investigations found that this was due to the size of the images. The images in the pilot study were based on the originals from the Radboud database and were therefore very large, high quality JPEG graphics which took quite some time to load. This had not been apparent when designing the images on a local network but as soon as participants were accessing the images as part of the survey and needing to download them from the Qualtrics servers it became apparent that this was taking several seconds per image and making the survey rather slow to complete. This could be quite off-putting to participants and might have caused them to abandon the survey part of the way

through. To address this, the images were resized to give a smaller file size but no appreciable loss in quality.

- Platform Recommendation: One goal of pilot testing was to have the survey completed using as many different platforms as possible. A concern was that given the recruitment strategy there may be a high proportion of participants trying to complete the survey on mobile devices so pilot testers were asked to try the survey out on their tablet devices and mobile phones as well as standard laptop and desktop computers. This was an extremely valuable test: firstly, it was found that the slider interface used in the pilot version of the survey was particularly cumbersome to use on iPad devices but also that the survey was impossible to complete using several older smart phones as the images were too large to view comfortably. By this stage the images had already been resized to address the loading speed issue detailed above and it was felt that any further reduction would begin to impact on the image quality when viewed on larger screens. Rather than try to accommodate all platforms with limited success, it was decided that a better approach would be to provide advice to participants on which platforms would be best to use to complete the survey. The initial instruction screen was modified to include advice to this effect.

Once these amendments and improvements were made and reviewed, the final version of the survey was created and the full set of FEBs loaded. At this stage it was ready for launch and recruitment was started.

Materials

The images in this study were drawn from the Radboud Faces Database (Langner et al, 2010). This is a database of high-quality face images where the models had been trained to consistently pose expressions as defined by the Facial Action Coding System. The

images in the database had been validated for reliability of emotional expression recognition plus mean intensity, clarity, genuineness and valence of the expressions shown. Permission was sought from the developers to use the database for this study, and a subset of the images were downloaded for manipulation, containing different emotions posed by individual models. Each set contained a happy and neutral image, plus angry, fearful, disgusted and sad images. These were chosen as representing core emotions that should be universally and clearly identifiable in their unmodified state (Ekman, 1992). As described above, the eye regions from selected faces were blended onto target faces displaying different emotions, producing a total of 24 image blends from each model. 48 models (19 female, 29 male) were used in this study: blends created from 24 models (12 female, 12 male) were used in Task One where inverted versions of their blends were also created by using a built-in rotation and flipping commands in the Pixelmator⁴ graphic editing package to rotate the images 180° and then flip the left-to-right orientation. 12 models were used in Task Two (6 female, 6 male) where inversion was not explored so no inverted versions were created. The remaining images from 3 sets of models (1 female, 2 male) were used in the study as training examples rather than experimental images.

Procedure

Participants were recruited using the methods described for previous studies, a mailing list used by post-graduate psychology researchers and two recruitment web sites⁵. In addition, a link to the study was distributed via social media, recruiting a large number

⁴ <http://www.pixelmator.com/mac/>

⁵ <http://psych.hanover.edu/research/exponnet.html>
<http://www.onlinepsychresearch.co.uk/researchers/>

of participants. The study was presented online using the Qualtrics⁶ system. The diagram below describes the structure of the experiment and the flow that participants took through each task.

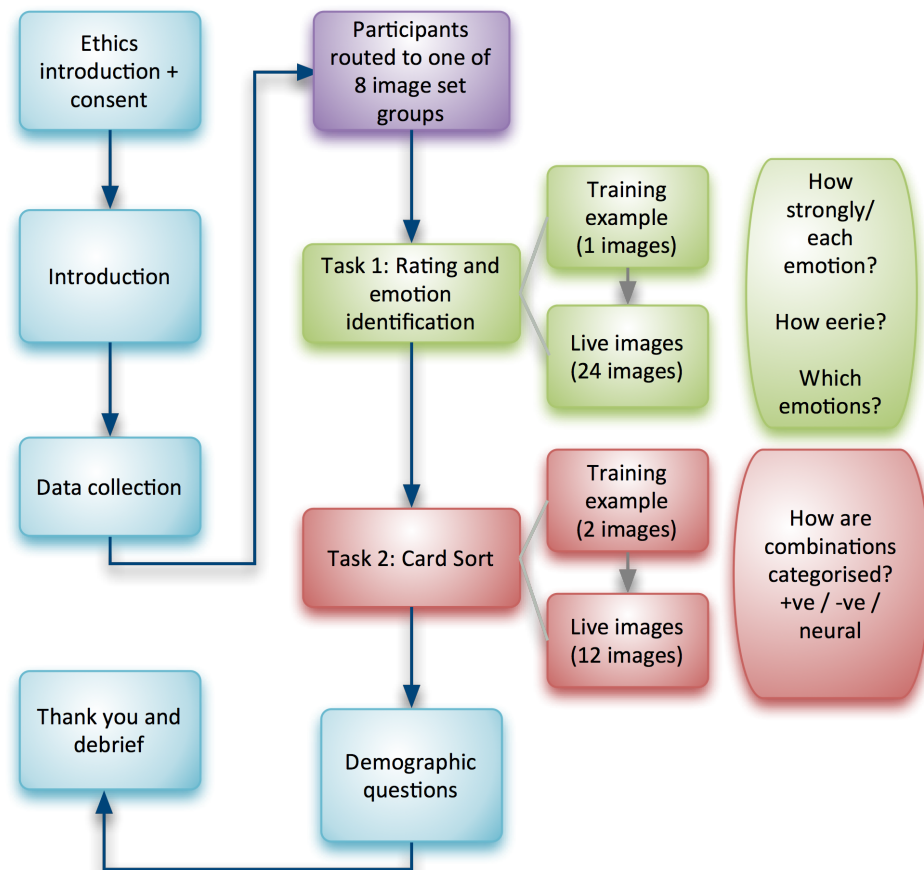


Figure 38: Diagram showing the survey flow and elements encountered by participants.

After they had completed the initial screens, participants were randomly allocated to one of eight groups. It was possible to randomly allocate participants to an experimental group and then present the elements within the study in a random order. The system presented a count of how far through the participant had progressed, displayed underneath the image that they were looking at at that moment in time. In this way, it

⁶ <http://www.qualtrics.com>

was clear how many images they still had to review. Also, if a question was not answered and the participant tried to progress to the next page, a pop-up appeared asking them if they were sure that they want to continue without answering. They were able to do so, but the pop-up encouraged them to go back and add their answers.

The structure of the experiment was as follows:

Introduction: The initial landing page introduced the study and the researcher and gave background information about the benefits and implications of participation. Participants could only proceed to take part in the study if they clicked to indicate that they were over eighteen years old and had understood the terms and conditions sufficiently to give their consent.

Task One pages: Once participants had given their consent to take part they started Task One. This began with a training phase where an example FEB was presented: this FEB was unique to the training phase and not used in any other part of the study. Participants were asked the three questions (strength of emotional response, perceived eeriness and emotion identification) about it that they would see in the live trials. Full details of the question texts are given in Appendix 3b. Each question page had the same format where the FEB image filled the top part of the visible area of the screen with the scale or grid presented below. This training example allowed participants to get used to the interface and also to check that the images were loading quickly and correctly for them. After the training phase, 24 FEBs were each presented and the three questions asked about them. The order of the FEBs was randomised. Participants could see how far they had progressed through the task as 'n/24' was presented under the image where n referred to the current FEB. When all 24 had been presented and participants had given their answers, they moved on to Task Two.

Task Two pages: Task Two again started with a practice example. The interface for Task 2 was quite different to a standard survey page so the training phase allowed participants to check that their equipment was capable of completing the task. For this task, the screen was split with an information area at the top and an interaction area in the centre of the screen. To the left of the interaction area was a 'stack' of images, and to the right a grid of four squares, each labelled with a category identifier. The participants' task was to pick up the images from the stack and place them into the category which they felt best represented that face. The training example only included two FEBs and again they were unique to the training part of the study. Once they had completed the training example they moved to a live version where there were 12 images to categorise. Once these had all been categorised, participants clicked to move to the last section of the study.

Demographic information: Finally, participants were asked if they were willing to give some basic demographic information. Only if they said that they were willing to give that information were they presented with these optional questions: they were age, gender and location as before but with the addition of ethnicity for this study only. Finally, participants were asked if they would be interested in receiving further information about the project including a report of the eventual results, and if so, they could leave a preferred email address.

Close: Once the final questions had been answered, participants were routed to the main project web page for further reading about the UVT. The landing page repeated the contact information for the researcher if there were any further queries.

Screenshots showing an example of each page in the structure are given in Appendix 3b.

Participants

3082 participants visited the first page of the survey and of those, 3077 gave their consent to start the study. 2316 participants completed the first training question and 1078 participants went on to complete both tasks. The analyses in the rest of this report are based on those participants. Of those, 999 of those were willing to give their demographic information; of those, 66% were female and the mean age was 37 years old. When asked about their location, 47% of participants reported that they were in the UK with another 35% reporting they were in the US. When asked about their ethnicity, most participants (56%) described themselves as white.

A full breakdown of participant demographic data is shown in Appendix 3a.

5.3: Results and Discussions

5.3.1: Eeriness ratings for different face image types

5.3.1.1: Results

The following analyses are drawn from the second question asked during Task 1 of Study 5. The idea that some faces may be perceived as eerier than others is key to the UVE, and this task evaluated responses to manipulated photographs of faces where emotional expressions from the eye and face region had been systematically combined to test two hypotheses about how these different combinations would affect whether participants would find them eerie. H1 predicted that the blended faces, where there was a mismatch between the emotions displayed in different parts of the face, would be rated as eerier than the original unchanged face images. This prediction was drawn from findings from earlier phases of the present project, and three potential patterns for the nature of this difference were explored, based on Tinwell et al's (2013) work linking the UVE to perceptions of psychopathy, Ekman et al's (2005) theory of leaked emotions, and

Seyama and Nagayama's (2009) findings that enlarged eyes triggered perceptions of eeriness. H2 predicted that orientation would have a significant effect on eeriness ratings.

Data preparation

To explore the hypotheses that mismatched faces would be found to be eerier than matched faces, and that this would be affected by inversion, the raw eeriness rating data was prepared for analysis. Participants had each seen 24 face images, half presented upright and half inverted. 4 sets of different models were used to create the images, so the raw results files were processed to give an eeriness rating score for each participant for each image, in the orientation they saw, regardless of the model used. Six of the 24 images were matched (the basic happy, neutral, angry, disgusted, fearful and sad faces) and the other 18 mismatched. This analysis stage was not concerned with which model had been used to create the specific combination of emotions, so the ratings from each participant were processed so that, regardless of the model they had seen, their eeriness ratings could be allocated to whether the face they had seen was an original, unblended 'matched' image, or one of the 'mismatched' blended faces. Responses were only analysed where participants had given ratings for matched and unmatched faces, so any responses where ratings were not given for both types of face were excluded.

The effect of inversion on ratings of eeriness for matched and mismatched faces.

This first test established whether there was a difference in eeriness ratings between the different face images, by analysing eeriness ratings to explore a possible interaction between match and orientation. A 2*2 within-participants ANOVA was used to examine the two main effects of match versus mismatch and orientation. The main effects of match and orientation were both significant: $F(1,1077) = 358.628, p < .001$, partial $\eta^2 = .25$ for the match-mismatch factor and $F(1,1077) = 367.105, p < .001$, partial $\eta^2 = .25$ for orientation. However, there was no significant interaction between the two factors:

$F(1,1077) = 4.918, p = .027$, partial $\eta^2 = .05$. The absence of an interaction is illustrated by the parallel lines in the figure below. The mismatched faces were rated as eerier than the matched faces, both when presented upright and when presented inverted. This provided support for H1 as the mismatched faces were rated as eerier than the matched faces, and in terms of H2 it indicated that inverting all the faces made them significantly eerier, regardless of whether they were matched or mismatched. This was the opposite to the predicted direction for H2, where it was expected that eeriness would be lower for the faces when presented inverted rather than upright.

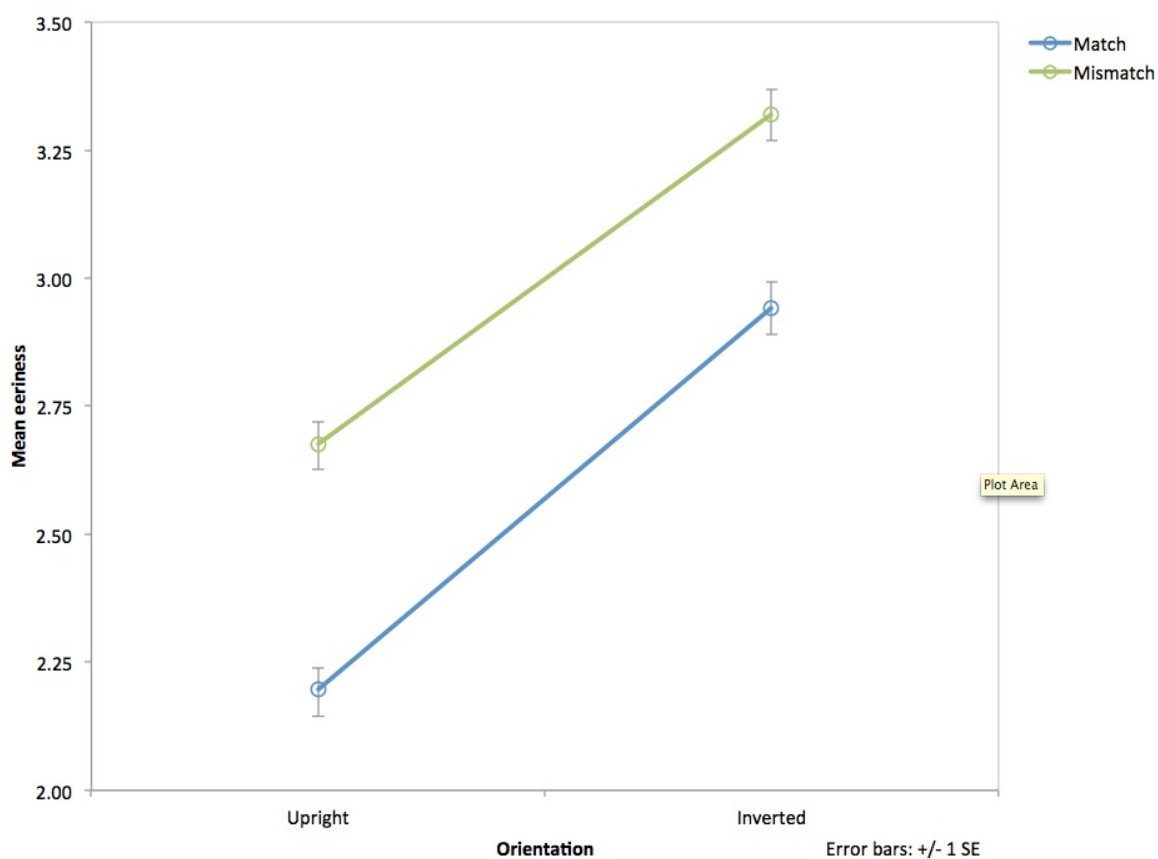


Figure 39: Comparison of the mean eeriness for matched and mismatched blends when presented upright and inverted.

This study was designed so that participants each saw a sub-set of face images drawn from a collection created from photographs of several different base models. This was to ensure that any observed effects could be applicable to faces in general rather than

grounded in a small set of specific images. However, this does mean that it was necessary to carry out an analysis by item to examine whether any of the variance in eeriness could be due to characteristics of the specific face images themselves, perhaps due to some expressions appearing particularly exaggerated or not clearly portrayed. To perform this, the original dataset of eeriness rating data was transposed to allow the 96 individual images (4 models x 24 face images) to be analysed by their match/mismatch status and the orientation in which they had been seen. A 2*2 ANOVA was carried out, which found the same results as the standard analysis by participant above as the main effects of match and orientation were both significant ($F(1,188) = 9.388, p < 0.001$, partial $\eta^2 = .048$), $F(1,188) = 24.038, p < 0.001$, partial $\eta^2 = .113$. There was no significant interaction between match and orientation: $F(1,188) = .183, p = .669$, partial $\eta^2 = .001$. As these results mirror those found in the standard analysis by participant, it was concluded that the images did not represent a confounding variable and could be considered reliable for the purposes of this study.

Which face images were rated as the most eerie?

The analyses above have demonstrated that the mismatched face images were perceived as eerier than the original unmodified face images, and this was found both when faces were presented upright and inverted. It is therefore valuable to consider in detail how the individual images differ from each other, to explore the specific predictions which had been made based on findings from the UV and face perception literature.

The first prediction was based on Tinwell et al's (2013) work which suggested that in general, the negative faces with the neutral eyes would be the eeriest, and of those, the disgusted, fearful and sad faces would be eeriest. The second prediction, based on Ekman et al's (2005) theory, was that the eeriest faces would be those with a happy face and negative (angry, disgusted, fearful or sad eyes). The final prediction, based on Seyama &

Nagayama (2009), suggested that the mismatched images with the largest eyes would be rated as the eeriest, which in this study would be the happy face with the fearful eyes.

The second hypothesis was that there would be a difference in eeriness as an effect of inversion, and the results of the ANOVA presented above indicated that inversion made all faces appear eerier. So, to explore the face types that were rated as particularly eerie, a heat map was constructed to visualise the variation in eeriness for each face image and orientation:

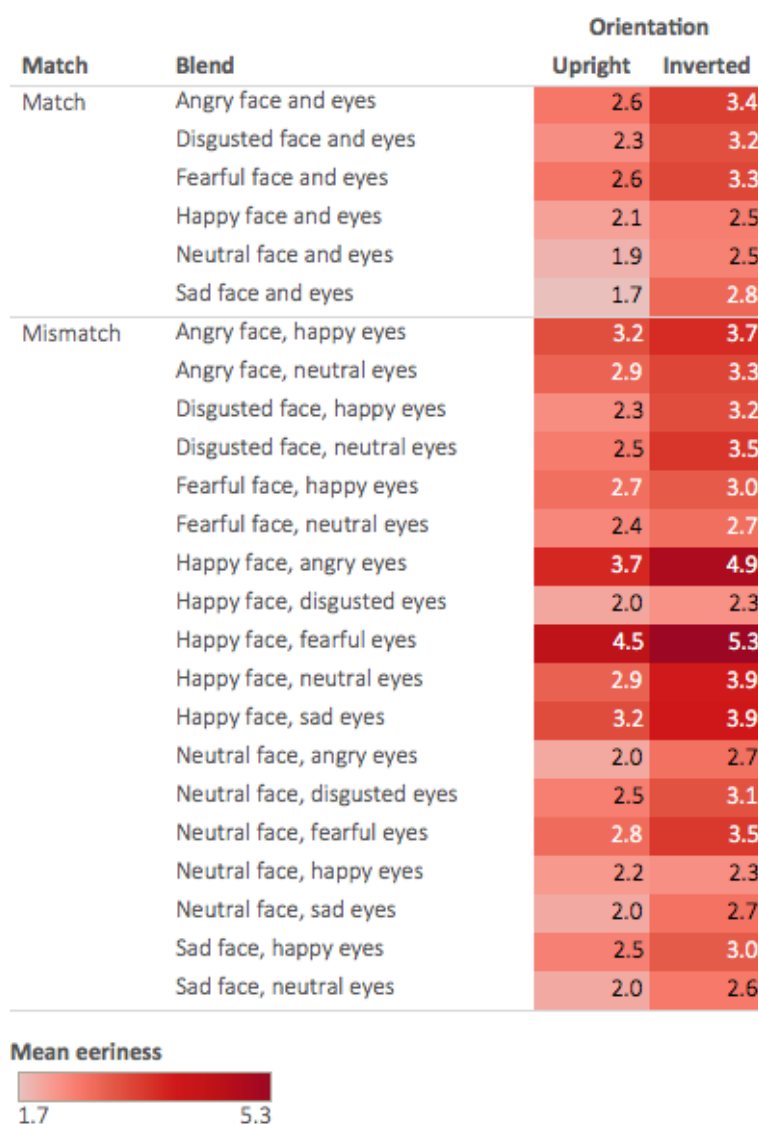


Figure 40: Heat map visualising eeriness ratings for each individual face image, presented upright and inverted.

It can be seen that the eeriest images were those where happy faces were blended with the negative emotions, with the *happy face, fearful eyes* and the *happy face, angry eyes* being the two eeriest images. In terms of the three hypotheses, these results start to indicate support for the predictions grounded in Ekman's (2005) work and Seyama and Nagayama (2009), but not for Tinwell (2013). Inferential analyses were then performed to see if these differences were statistically significant.

Two 3*2 mixed ANOVA tests were used to analyse whether there was a significant difference in eeriness between the blends containing fearful and angry eyes and the basic unblended fearful and angry faces. Firstly, the *happy face, fearful eyes* blend, the *neutral face, fearful eyes* blend and the basic fearful un-blended face were compared. Examples of these three faces are shown in Figure 41 below:

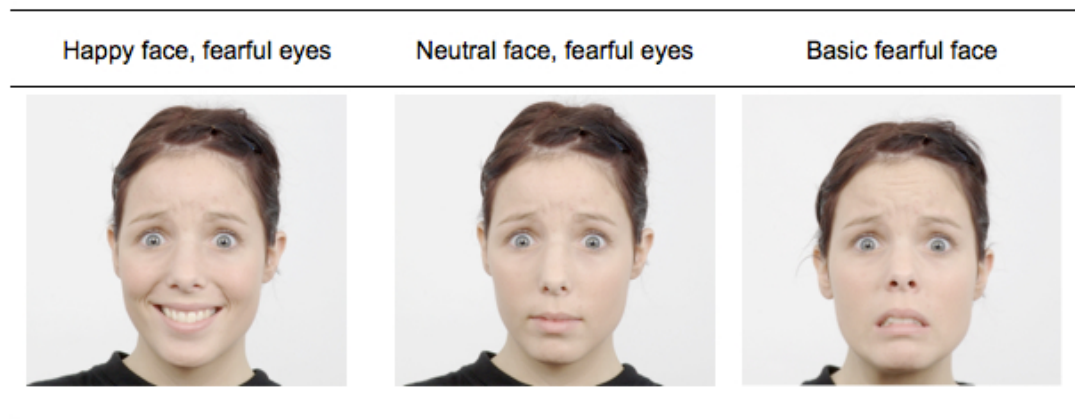


Figure 41: Examples of the three face images where fearful eyes were presented, either blended with a happy or neutral face or in the unblended fearful face.

This test analysed two factors, the within-subject factor of face image with three levels (*happy face, fearful eyes; neutral face, fearful eyes; and unblended fearful face*). The second, between-subjects factor was inversion, with two levels, upright and inverted. As these analyses served as post-hoc tests, exploring the main effects found in the ANOVA above, Bonferroni's correction was applied to the target p value for the three comparisons, so significance was assessed at $p < 0.017$. The main effect of blend was

significant: $F(2,2152) = 398.268, p < .001$, partial $\eta^2 = .27$. The main effect of orientation was also significant: $F(1,1076) = 3934.13, p < .001$, partial $\eta^2 = .785$, with eeriness ratings higher when the images were presented inverted compared to upright. The interaction between face image and orientation was not significant: $F(2,2152) = .389, p = .678$, partial $\eta^2 = .0$. Post-hoc pairwise comparisons, again using Bonferroni's adjustment at $p < 0.017$ for the comparisons indicated that there were significant differences between two of the images: the differences between the *happy face, fearful eyes* and *neutral face, fearful eyes* blends and the *happy face, fearful eyes* and *unblended fearful face* images were significant but the *neutral face, fearful eyes* blend was not significantly different from the *unblended fearful face* ($p=0.77$). The figure below illustrates the mean eeriness for each orientation and face image.

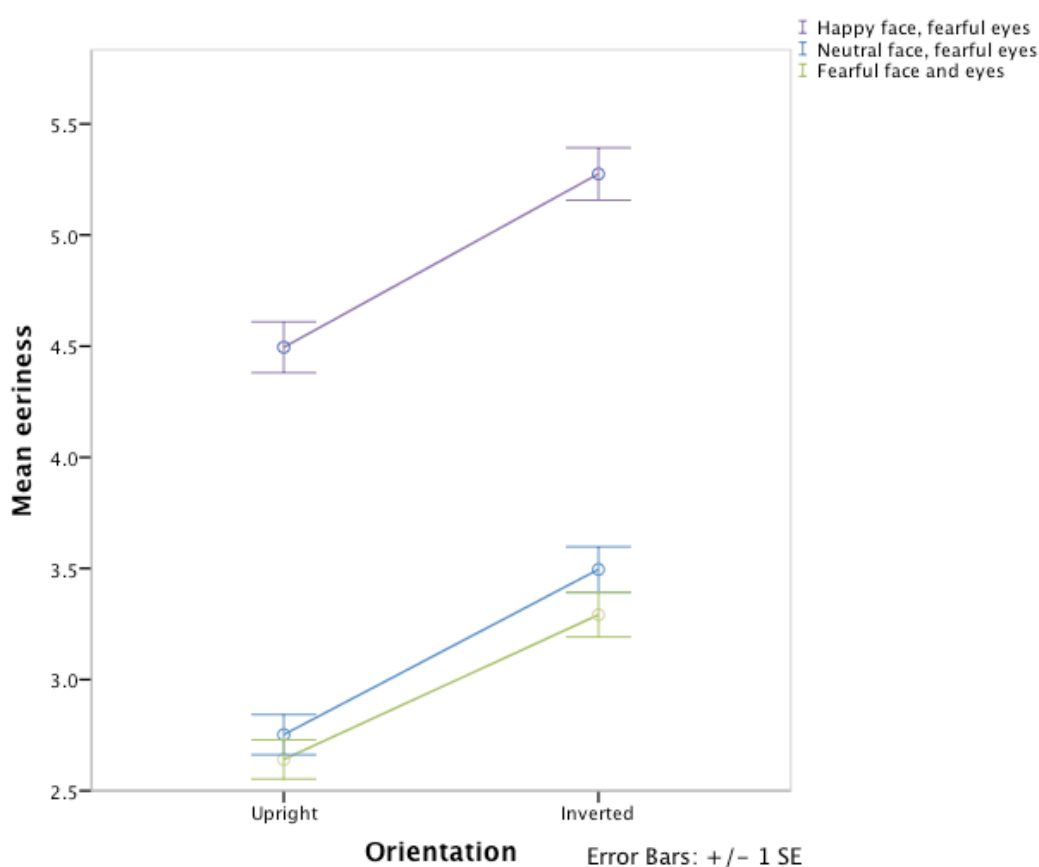


Figure 42: Comparison of mean eeriness for fearful-eyed faces when presented upright and inverted.

These results can be used to consider the third prediction for the nature of H1, based on the findings of Seyama and Nagayama (2009), and it can be seen that this was not supported as the eeriness ratings were not highest for the *neutral face, fearful eyes* blend, nor were they significantly different from the unblended fearful face. However, they begin to provide support for H2, grounded in Ekman's (2005) research, which proposed that the faces images where the negative eyes were blended with happy faces would be rated as most eerie.

A second 3*2 mixed ANOVA was performed to analyse the second group of faces which seemed particularly eerie in the heat map visualisation. These were the face images including angry eyes: *happy face, angry eyes* blend, the *neutral face, angry eyes* blend and the basic angry un-blended face. Examples of these three faces are shown in the Figure below:

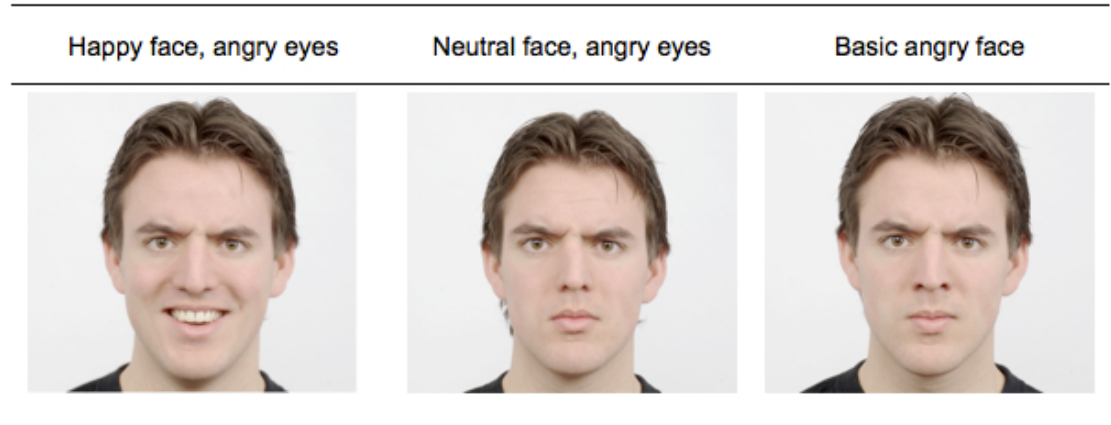


Figure 43: Examples of the three face images where angry eyes were presented, either blended with a happy or neutral face or in the unblended angry face.

As with the fearful-eyed faces, the main effect of face image was significant: $F(2,2152) = 350.924$, $p < .001$, partial $\eta^2 = .246$ and the main effect of orientation was also significant: $F(1,1076) = 3630.345$, $p < .001$, partial $\eta^2 = .771$. The interaction between face image and orientation was also significant when assessed at $p < 0.017$: $F(2,2152) = 5.762$, $p = .003$, partial $\eta^2 = .005$. Ratings of eeriness were higher when the images were

presented inverted rather than upright, with a more marked effect of inversion on the *happy face, angry eyes* blend. Post-hoc pairwise comparisons using Bonferroni's adjustment indicated that there were significant differences between all three images at $p < 0.017$. The figure below illustrates the mean eeriness for each orientation and image:

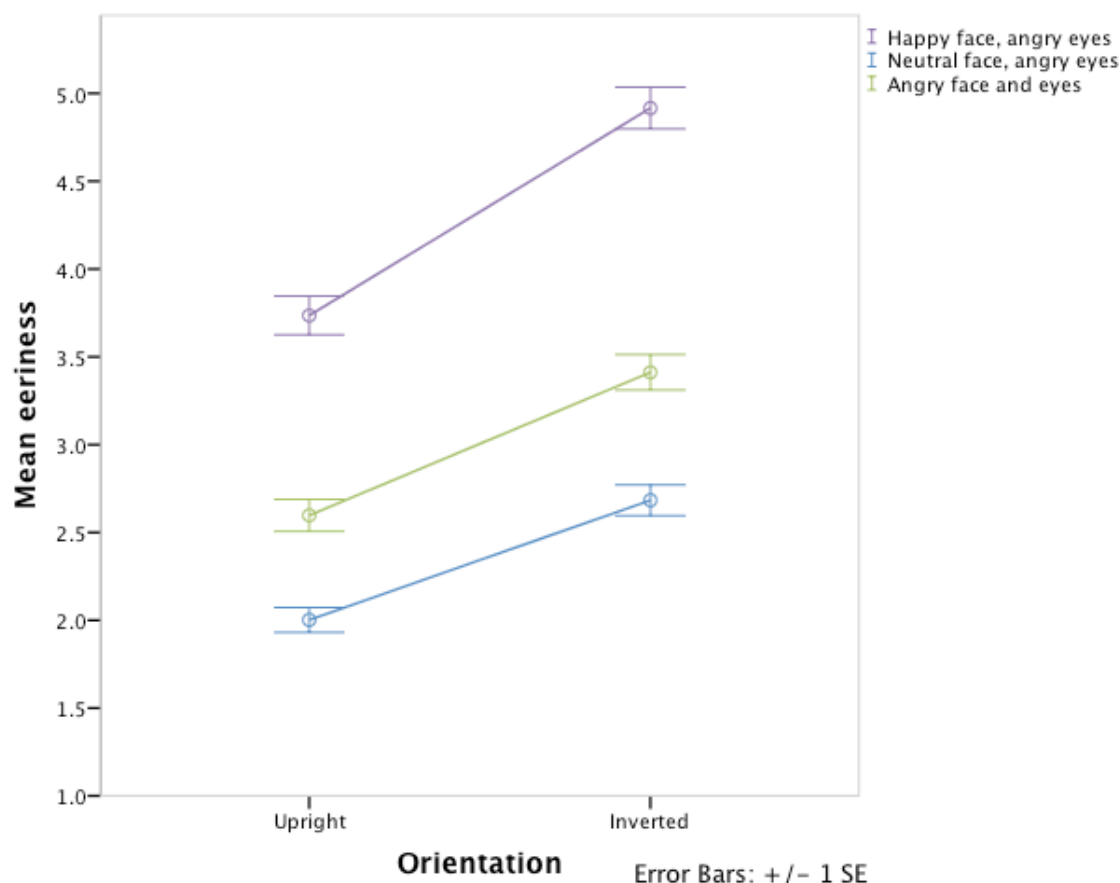


Figure 44: Comparison of mean eeriness for angry-eyed faces when presented upright and inverted.

These results mirror those found for the fearful blends, which does provide support for the prediction grounded in Ekman et al's theory of leaked expressions which suggested that in general, the eeriest blends will be those where positive faces and negative eyes are combined. Post-hoc analysis of whether this is a general pattern across the image types was also considered. The blended faces images used in this study were created by taking posed emotional expressions and transforming them into blended faces by merging the

eye regions from one expression into the face of another. The general effects of matched and mismatched faces have been described above, and the two eeriest types of combination have been explored. Those two blends were both combinations of a positive face with negative eyes, but would a more general pattern be found across all of the blends where negative eyes had been blended with a positive face? One explanation for the mismatch effect found in H1 proposes the principle that these would be the eeriest faces. To explore this, the individual face images were combined into groups for analysis, where the 24 individual images were grouped into nine combinations of positive, negative or neutral faces and eyes. For example, the *happy face, fearful eyes* blend is an example of a *positive face, negative eyes* blend type and the *sad face, happy eyes* blend an example of a *negative face, positive eyes* blend type. By grouping the faces in this way, it is possible to move beyond examining individual face images and to test the patterns of eeriness between these categorised types. Examples of the nine face image types are shown in the table below.

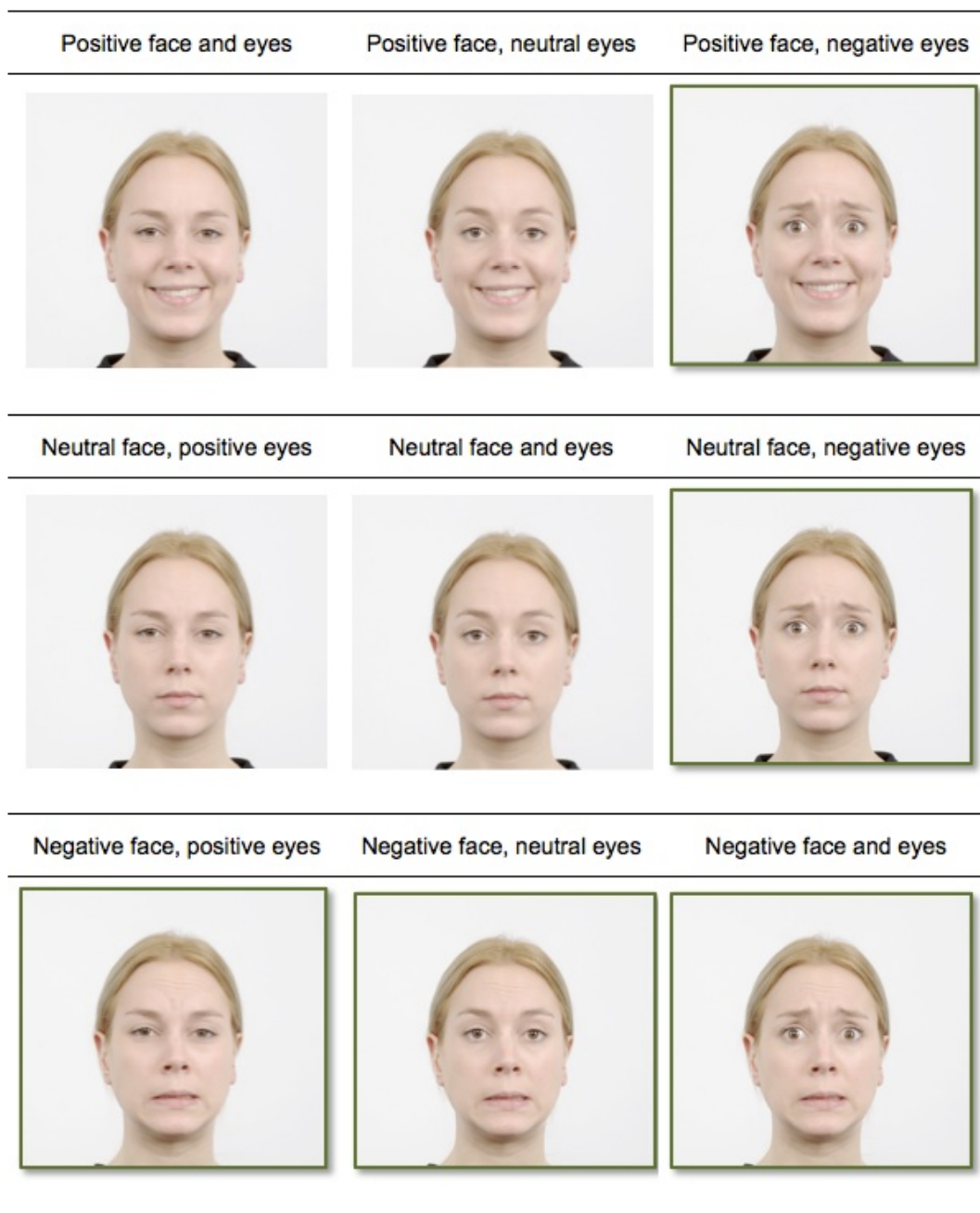


Figure 45: Examples of the positive, negative and neutral face image types

As each participant saw half of the blends upright and half inverted, a new variable for image type in each orientation was calculated from the eeriness ratings for the individual blends. For example, the upright *positive face, negative eyes* blend type value for each participant was calculated as a mean of the ratings they gave for the two of the four blends (*positive face/angry eyes*, *positive face/disgusted eyes*, *positive face/fearful eyes*, *positive face/sad eyes*) that they saw in the upright orientation. This variable could then

be analysed to give the mean eeriness rating by face image and orientation as shown in Figure 46 below. The eeriest blend type was the *positive face, negative eyes* type, and the inverted images were again rated as eerier than the upright blends.

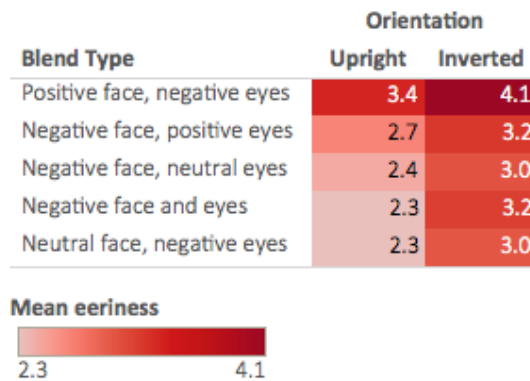


Figure 46: Mean eeriness ratings for positive, negative and neutral blend types, by orientation.

To test whether the *positive face, negative eyes* blend types would be found to be significantly eerier than the other four types, the eeriness ratings for the upright blend types were analysed using a within-participants ANOVA. Mauchley's test indicated that the assumptions of sphericity had been violated: $\chi^2(9)=223.67$, $p < .0005$, and so degrees of freedom were corrected using the Greenhouse-Geisser estimates of sphericity ($\epsilon=.90$). The corrected results show that blend type did have a significant effect on eeriness: $F(3.588,3863.90) = 147.629$, $p < 0.005$, partial $\eta^2 = .121$. Post-hoc pairwise comparisons using Bonferroni's adjustment indicated that there were significant differences between three of the five blend types at $p < 0.013$: significant differences were found between all the blends except for the *negative face, neutral eyes blend type* and the *neutral face, negative eyes blend type*, or between the *neutral face, negative eyes blend type* and the *negative face and eyes blend type*. The figure below illustrates the mean eeriness for each blend type.

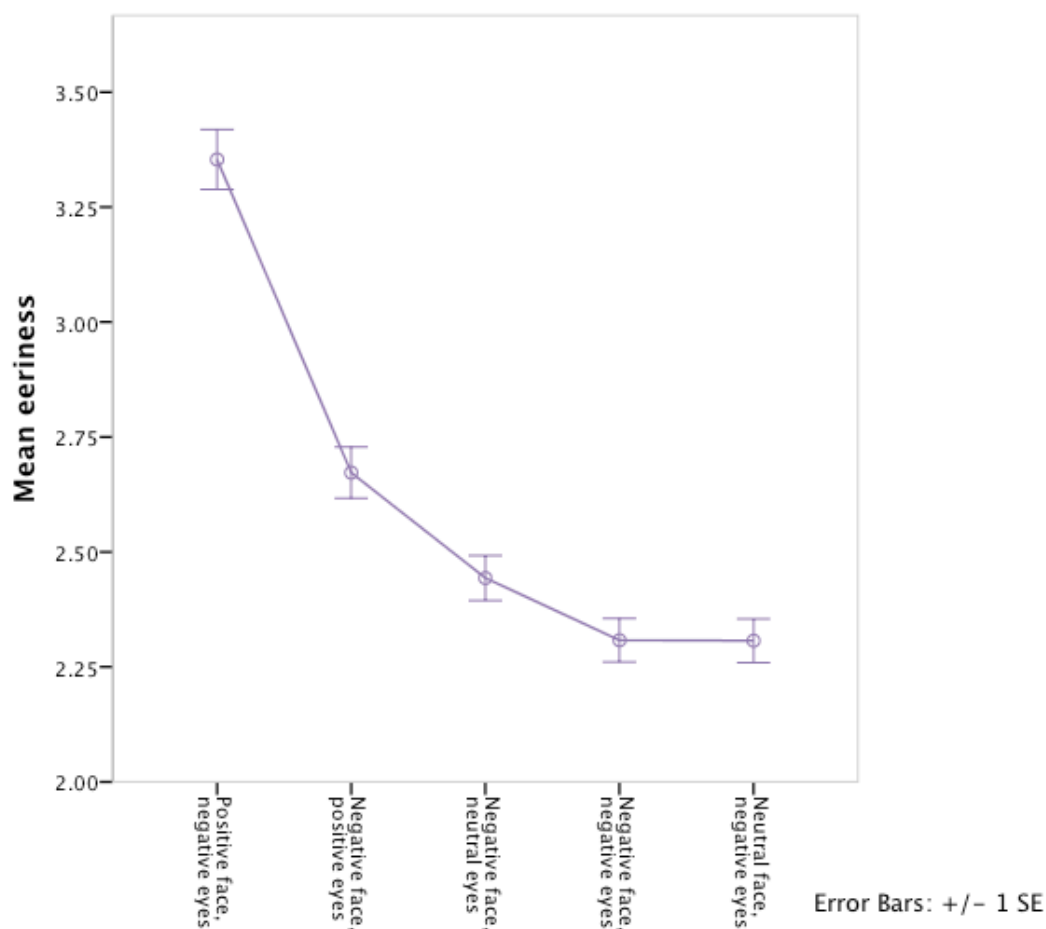


Figure 47: Comparing mean eeriness ratings for face image types when presented upright.

These results provide support for Ekman et al's prediction as the type of combination where negative eyes paired with a positive face were rated as the eeriest type.

5.3.1.2: Discussion

Having presented the results of the eeriness analysis, the implications of these findings will now be briefly described before the results of the emotion response data are presented. The core research aim for the analysis of the eeriness data was to establish whether mismatched faces would be perceived as eerier than matched faces. This hypothesis was supported, as mismatched faces were found to be rated as significantly more eerie than the matched faces. This was found at the overall level and also when

blend types were compared. The blend types that was rated as the most eerie were the ones with the greatest 'distance' of a mismatch between the eyes and face, the *positive face, negative eyes* face and *negative face, positive eyes* face. At the level of individual images, the least eerie was an original, unblended image, and this was the original sad face which had a mean eeriness rating of only 1.7. These significant findings will now be considered in the context of the theoretical background to the UVE and also to face perception in general.

The UV literature suggested two potential patterns for the subjective ratings of eeriness that may be found in this study, based on Tinwell et al's (2013) theory of perceptions of psychopathy and Seyama and Nagayama's (2009) findings that enlarged and exaggerated eyes were key to evoking a sense of the uncanny. In addition, Ekman et al's (2005) theory of leaked expressions was applied to the idea of the UVE, even though this theory pre-dates the idea of an UVE.

There was the most support for the pattern of response described by Ekman et al, as leaked emotion theory would propose that the most aversive face blends would be those with a marked incongruity between a smiling face and negative expression which was 'leaked' in the expression conveyed by the eyes. The blends in the present study would be those blends with a happy face and the angry, afraid, sad or disgusted eyes. The blend type analysis found that those *positive face, negative eye blends* were the most eerie and when blends were individually considered, the ones found to be most eerie were those with the *fearful* and *angry* eyes in the happy faces. In terms of leaked emotion theory, these "poorly concealed" emotions would cause people to be negatively evaluated because it makes them appear untrustworthy. The findings of the present study may complement this general theory by suggesting that some leaked emotions are more aversive than others. The happy faces with angry and fearful eyes were found to be the

most eerie and this may be due to the fact that these actors were perceived as attempting to mask their genuine feelings of anger or fear, as someone who is pretending not to be angry when they actually do feel that emotion strongly may be dangerous or unpredictable. A masked expression of fear may elicit concern as to why this concealment is being attempted. Of the four emotions used in this study, combinations of anger and frightened eyes in a happy face are suggestive of scenarios which could be read as potentially threatening while someone attempting to conceal feelings of sadness or disgust are less likely to represent imminent danger. It may actually be that these blends were also less eerie because they mimic expressions that are more commonly experienced in social encounters where everyday conventions of politeness mean that it is quite normal to cover up one's sense of sorrow or revulsion.

The hypothesis based on Tinwell et al's (2013) theory, that the blends where the neutral eyes were presented in negative-expression faces would be rated as the most eerie, was not supported. Anecdotally, NHAs proposed as inhabitants of the uncanny valley are often described as 'dead-eyed'. In addition, the results of the studies described in Phase 1 of the present thesis found that the NHA with large, blank eyes was found to be most eerie and aversive. However, in this study the neutral eye blends were only moderately eerie as the *negative face, neutral eyes* blend type was only the third eeriest, with a mean eeriness of 2.4, and the *positive face, neutral eyes* blend itself only had a mean eeriness of 2.9.

Based on Tinwell et al's (2013) theory, it was possible to predict patterns of eeriness for the individual blends when the four different negative-emotion faces were combined with neutral eyes and the disgusted, fearful and sad faces were hypothesised to be the eeriest within this subset of blends. When analysed, it was found that angry and happy faces with neutral eyes were the eeriest, as shown in the figure below.

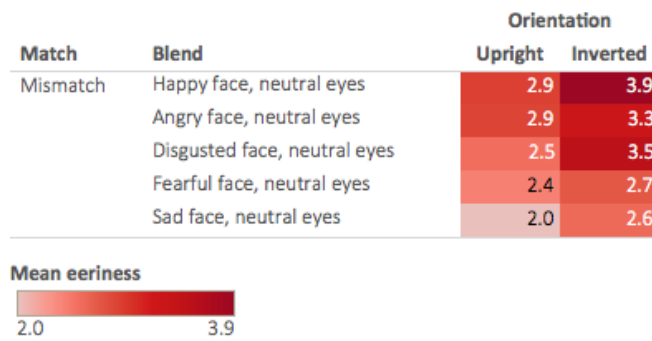


Figure 48: Heat map visualising eeriness ratings for the five neutral-eyed blends.

While the ‘dead-eyed’ theory may not have been supported by these results, the finding that the eeriest faces were those where the expressions were mismatched suggests that an element of non-congruence may be at the heart of the UVE but in a different manner to that found by Tinwell et al as their type of mismatch involved inexpressive eyes but the results in this study suggest a particular type of positive-negative mismatch rather than a general lack of expression in the upper part of the face. These results are supportive of Creed and Beale (2008) which suggested an incongruence in emotional expression may be responsible for the UVE.

The final scenario for consideration here is that based on Seyama and Nagayama (2009) study which found that enlarged eyes were a factor in triggering a sense of the uncanny in artificial faces. The blended faces in this study included some where the eyes appeared large, even exaggeratedly so, where frightened eyes were blended with happy or neutral expressions. These blends showed eyes widened in fright whereas the angry, disgusted and sad expressions would not be characterised by a widening of the eyes. Therefore, the blends with fearful eyes will be rated as the most eerie and of those, the neutral face with the fearful expression will be rated as eeriest. It was found that the first part of this hypothesis was supported as the eeriest blend was one with fearful eyes but it was the *happy face with fearful eyes* rather than the *neutral face with fearful eyes* that was rated as eeriest.

Three scenarios relating specifically to theories of the UVE have been considered, but the eeriness findings can also be discussed more broadly in terms of implications for face perception by examining the pattern of these eeriness ratings when the blends were presented inverted rather than upright. This was considered in H2, and the results suggest two notable observations when orientation is considered as a factor for analysis. Firstly, significant differences in eeriness were found for mismatched and matched faces in either orientation. Secondly, it was found that all of the individual blends were rated as eerier when presented inverted compared to when they were presented upright. Figure 49 below presents the difference between the upright and inverted eeriness ratings, coded to highlight the blends where the inverted faces were rated as eerier than the upright faces.

| Match | Blend | Orientation | | Difference |
|----------|------------------------------|-------------|----------|------------|
| | | Upright | Inverted | |
| Mismatch | Happy face, fearful eyes | 4.5 | 5.3 | 0.8 |
| | Happy face, angry eyes | 3.7 | 4.9 | 1.2 |
| | Happy face, sad eyes | 3.2 | 3.9 | 0.7 |
| | Happy face, neutral eyes | 2.9 | 3.9 | 1.0 |
| | Angry face, happy eyes | 3.2 | 3.7 | 0.5 |
| | Disgusted face, neutral eyes | 2.5 | 3.5 | 1.0 |
| | Neutral face, fearful eyes | 2.8 | 3.5 | 0.7 |
| | Angry face, neutral eyes | 2.9 | 3.3 | 0.4 |
| | Disgusted face, happy eyes | 2.3 | 3.2 | 0.9 |
| | Neutral face, disgusted eyes | 2.5 | 3.1 | 0.6 |
| | Fearful face, happy eyes | 2.7 | 3.0 | 0.3 |
| | Sad face, happy eyes | 2.5 | 3.0 | 0.5 |
| | Fearful face, neutral eyes | 2.4 | 2.7 | 0.3 |
| | Neutral face, angry eyes | 2.0 | 2.7 | 0.7 |
| | Neutral face, sad eyes | 2.0 | 2.7 | 0.7 |
| | Sad face, neutral eyes | 2.0 | 2.6 | 0.6 |
| | Neutral face and eyes | 1.9 | 2.5 | 0.6 |
| | Neutral face, happy eyes | 2.2 | 2.3 | 0.1 |
| | Happy face, disgusted eyes | 2.0 | 2.3 | 0.2 |
| Match | Angry face and eyes | 2.6 | 3.4 | 0.8 |
| | Fearful face and eyes | 2.6 | 3.3 | 0.7 |
| | Disgusted face and eyes | 2.3 | 3.2 | 0.8 |
| | Sad face and eyes | 1.7 | 2.8 | 1.1 |
| | Happy face and eyes | 2.1 | 2.5 | 0.4 |

Figure 49: Face blends sorted by descending eeriness when inverted, highlighted to present difference between inverted and upright eeriness ratings.

These findings suggest support for the H2, following on from Study 4 detailed above, that blended faces may be processed differently to non-blended faces. It can be seen that the mismatched blends are rated as eerier than the matched blends, and that the eeriest inverted blends are again those with the *happy face, negative eyes* blends. The high ratings of eeriness for the inverted faces is notable as it does not follow the expected pattern of emotional response to inverted faces as generally, emotional expressions are harder to perceive from inverted faces (Prkachin, 2003; Fallshore and Bartholow, 2003). A close analogue to the blended faces used in this study would be Thatcherised faces (Thompson, 1980) where internal features are transformed by inversion, creating a

chimeric face which is grotesque in appearance when presented upright but does not appear grotesque when inverted. This is generally seen as evidence that faces are processed holistically as the inversion of the face disrupts the perception of the relationship between the individual features, making it harder to recognise that those features have been manipulated. Cornes et al (2011) found that ratings of Thatcherised faces when presented inverted were rated as considerably less grotesque than when presented upright. Applying that logic to the faces in the present study, it would be expected that inverted faces would be rated as *less* eerie than the upright faces. However, this was not the case which suggests that some analytic aspects of face perception are being engaged with these blended faces. This surprising finding does suggest that these results could be explored in more detail in future studies to find out why this pattern of responses occurred. One possibility may be that the nature of the task, in asking participants to focus on eeriness and to rate how eerie they found each face, may have primed them to be more sensitive to any atypical or unusual images, and in the course of day-to-day life, it is unusual to encounter an upside-down face. If this was the case, then the inverted versions of the images may have been rated as eerier because the orientation was unusual in itself, with the unsettling nature of the mismatched faces acting as a secondary influence to the rating. There were marked differences between the eeriness ratings for each of the face types, so although the higher eeriness of inverted faces is unusual, a lack of familiarity with inversion cannot be the only factor at work in causing this effect.

A second notable finding is that the pattern of difference between the upright and inverted face images is the same for the matched and mismatched faces. The mismatched faces were indeed found to be the eeriest overall, but it may have been expected that there would be a more marked effect of inversion for the original, unblended faces. It

would be expected that perception of natural faces would be more disrupted by the act of inversion. The fact that the same pattern of difference was observed for the original and blended images does suggest that inversion alone is not the sole reason for the variation in eeriness between these face images. It may be that as participants were explicitly asked to rate eeriness in the images, some form of analytic processing was being engaged for all of the images, regardless of whether they were matched or not. The implications of these findings will be considered in more detail in Section 5.4, the general discussion of the findings in this chapter.

5.3.2: Emotional responses to face image types

5.3.2.1: Results

How do feelings of eeriness relate to those of fear, anger, sadness, disgust, happiness and surprise?

The following analyses are drawn from the first question asked during Task 1 of Study 5. Participants gave ratings of how strongly they felt six emotions: anger, fear, happiness, disgust, sadness and surprise, on a scale of 1-9 scale, with 1 labelled as 'I don't feel this at all', 5 labelled as 'I feel this moderately strongly' and 9 labelled as 'I feel this extremely strongly'. (See Table 12 for the full question wording and response labels.)

These responses were gathered to explore the relationship between eeriness and other emotions, with a view to understanding more about the emotional composition of the UVE. Eeriness is seen as the key emotion characterising the effect but the relationship between it and other, more basic, emotions has not yet been explored systematically for NHAs. By mapping the relationship between these ratings for these blended faces it is possible to test two hypotheses. Firstly, H3, was based on Ho et al (2008) who reported that eeriness for NHAs should be strongly associated with feelings of fear and disgust, so

the mismatched blends would also evoke the strongest feelings of fear and disgust, and H4, based on Bassili (1979) which suggests that participant ratings of happiness and disgust will be higher when those emotions are presented in the eyes of the blend, and the ratings of anger and fear highest when these are the emotions presented in the rest of the face. To explore the nature of the emotional response ratings for each face blend, the mean ratings given by each participant were analysed and the heat maps in Figure 50 below present the mean emotion strength for each face blend on the six emotion scales. These results will be considered in detail below, but an initial reading of these maps indicated that the most strongly felt emotions were consistent with the expressions posed in the images and that emotions were generally felt more strongly when looking at the upright blends.

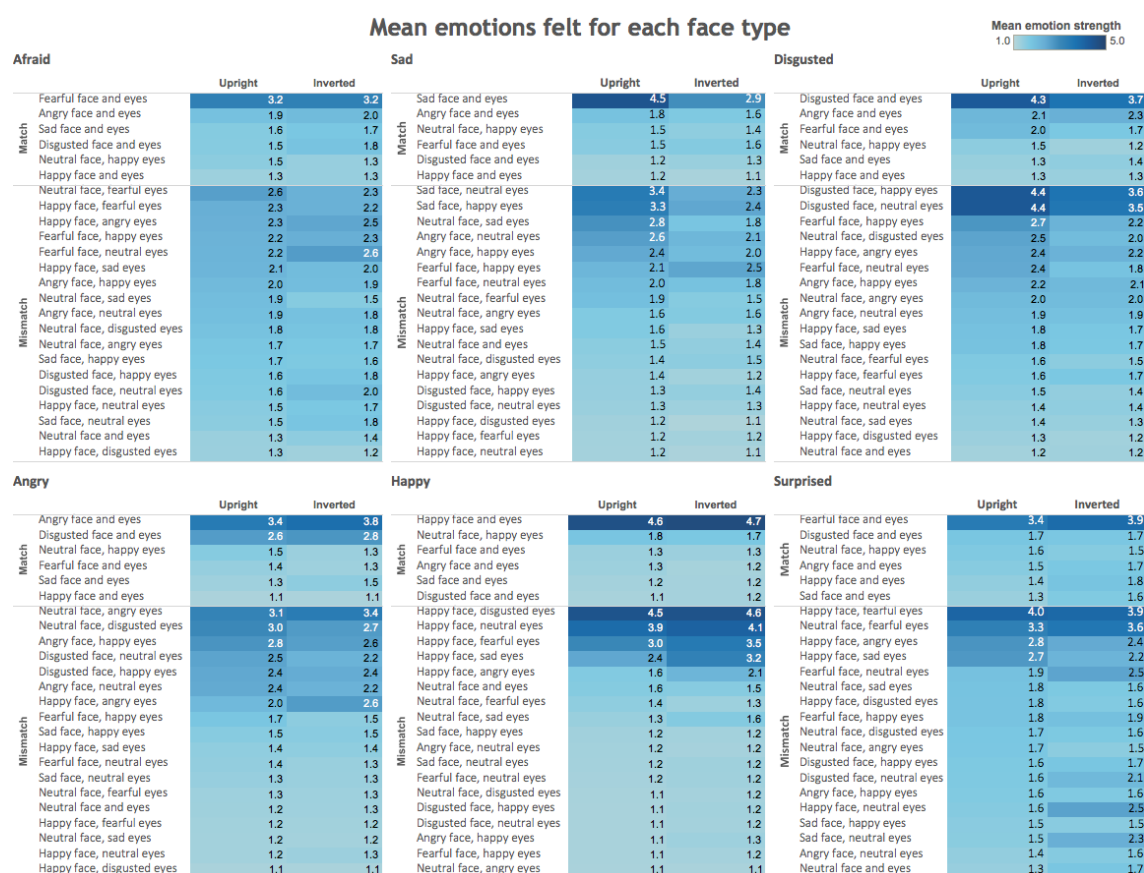


Figure 50: Mean emotion strength across all face blends, presented upright and inverted. Tables are grouped by whether the blend was matched or mismatched, and sorted in descending order of the emotion strength when presented upright.

To explore the nature of these differences, and test H3 above, four of these emotion scales were selected for detailed analysis. As the UVE is characterised as an unsettling or creepy feeling, feeling afraid is key to understanding the effect. It has also been suggested that the UVE has an aversive component so the ratings of disgust and sadness were also analysed. Happiness was analysed as a check against the negative emotions, as it would not be expected that blends evoking a strongly uncanny response would also make participants report feeling happy. Four 2*2 within-participants ANOVAs were carried out to examine these effects, and significant main effects and interactions were found in all cases.

The effects of face mismatch and orientation on feeling afraid

A 2*2 within-participants ANOVA was used to examine the two main effects of match versus mismatch and orientation on participant's reported experience of feeling afraid. The main effects of match and orientation were both significant: $F(1,1077) = 355.456, p < .001$, partial $\eta^2 = .248$ for the match-mismatch factor and $F(1,1077) = 71.593, p < .001$, partial $\eta^2 = .062$ for orientation. There was also a significant interaction between match and orientation: $F(1,1077) = 207.167, p < .001$, partial $\eta^2 = .161$. Considering the relative effect sizes for each of the reported fear ratings, it can be seen that the largest was that for the match versus mismatch factor, with the interaction between the factors having a medium effect size and the orientation itself only a small effect size. The interaction

between match and orientation is shown in Figure 51 below:

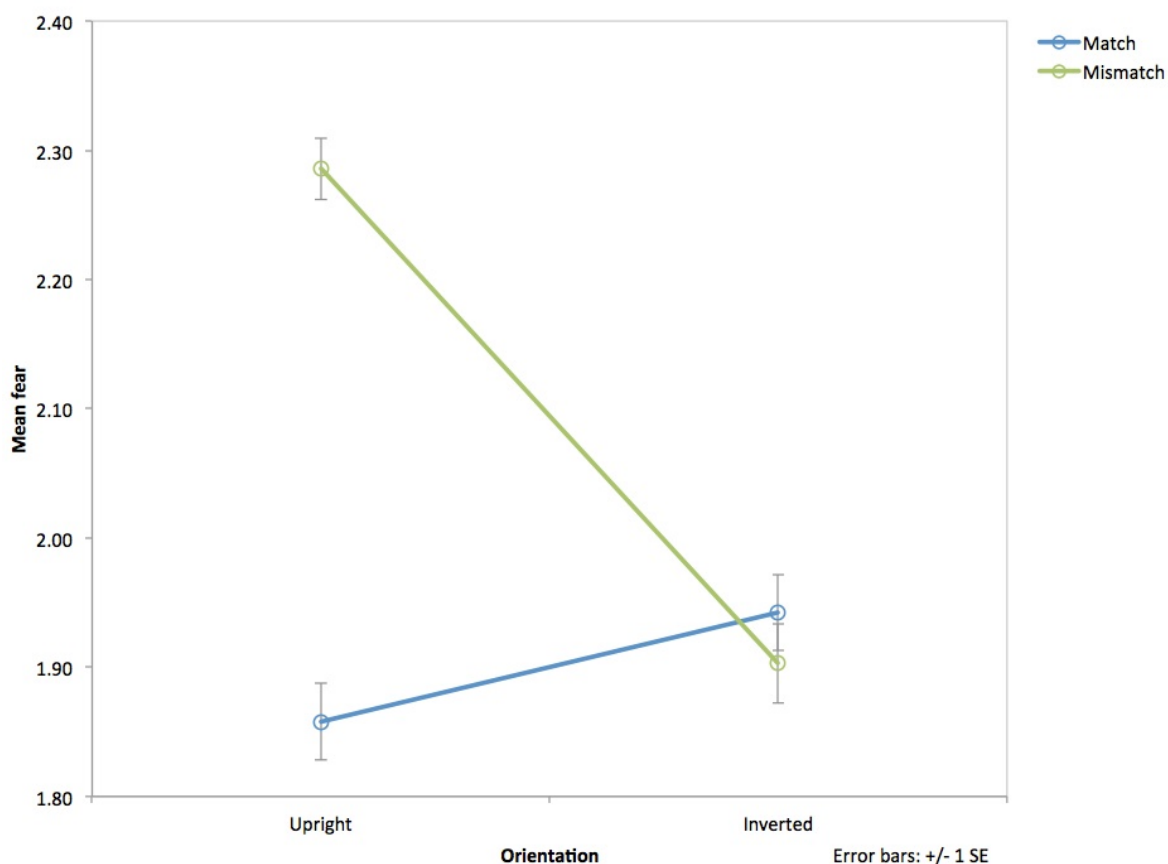


Figure 51: Comparison of the mean fear strength for matched and mismatched blends when presented upright and inverted.

Post-hoc t-tests were used to explore the nature of this interaction by testing the significance of each of the four combinations of match and orientation. The results of these tests are shown in Table 13 below.

| Comparison | <i>t</i> | <i>p</i> |
|--|----------|----------|
| Mismatched images only: upright vs inverted | -.889 | .374 |
| Matched images only: upright vs inverted | -3.727 | .000 * |
| Upright presentation only: mismatched vs matched images | 3.380 | .001 * |
| Inverted presentation only: mismatched vs matched images | -1.228 | .220 |

* = Significant at $p < 0.0125$

Table 13: Results of post-hoc t-tests on mean fear ratings

p values were assessed at <0.0125 , and it was found that matched images were significantly more frightening when presented inverted compared to upright and that

mismatched images were significantly scarier than matched images, but only when presented upright. When considering mismatched images only, orientation did not significantly affect fear ratings nor did mismatch versus match when only inverted images were considered. Given that the emotions of fear and eeriness may be considered similar, this finding is interesting when compared to the pattern of findings for the eeriness ratings, as these were higher for inverted presentations regardless of whether the images were matched or mismatched.

The effects of face mismatch and orientation on feeling sad

A 2*2 within-participants ANOVA was used to examine the two main effects of match versus mismatch and orientation on participant's reported experience of feeling sad. The main effects of match and orientation were both significant: $F(1,1077) = 436.911, p < .001$, partial $\eta^2 = .289$ for the match-mismatch factor and $F(1,1077) = 765.179, p < .001$, partial $\eta^2 = .415$ for orientation. There was also a significant interaction between match and orientation: $F(1,1077) = 378.538, p < .001$, partial $\eta^2 = .260$. All effect sizes were large in this analysis, but the largest was found for the orientation factor, suggesting it was most responsible for the variance in sadness ratings across these two factors. The interaction between match and orientation is shown in Figure 52 below:

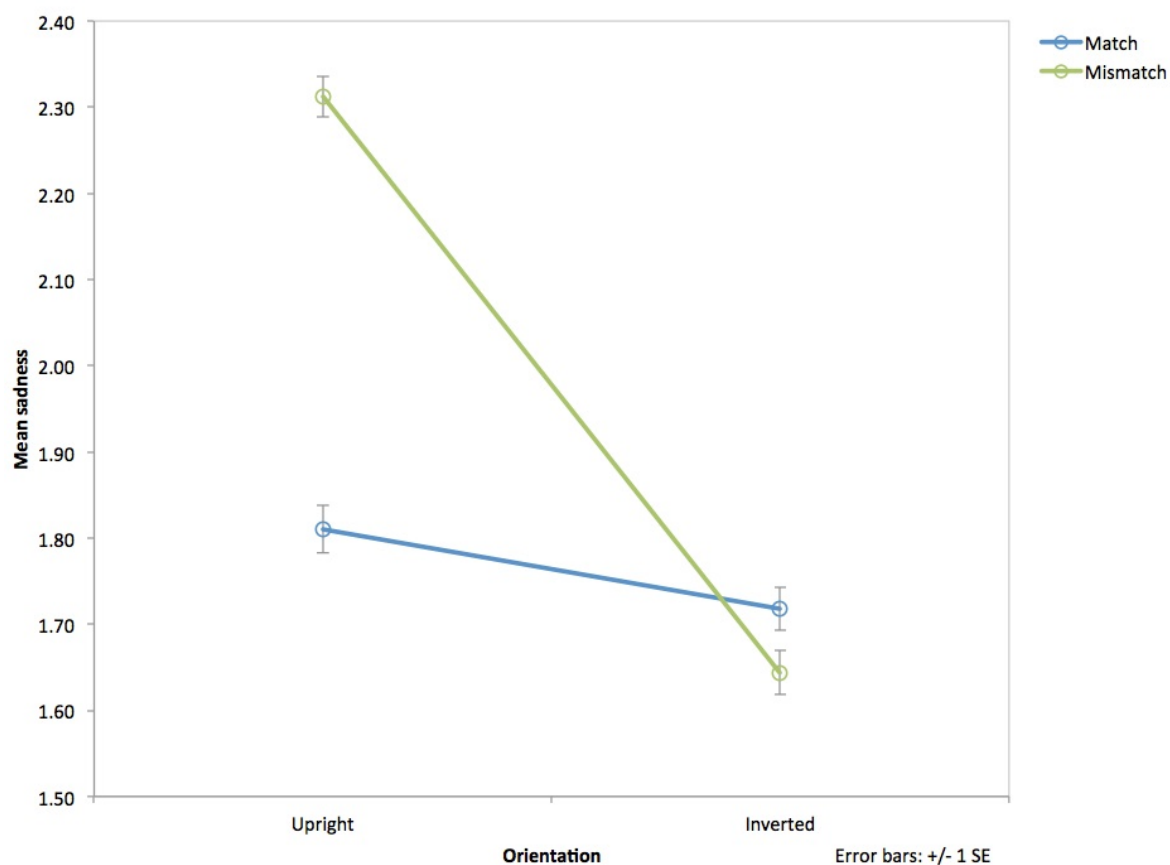


Figure 52: Comparison of the mean sadness strength for matched and mismatched face images when presented upright and inverted.

Post-hoc t-tests were used to explore the nature of this interaction by testing the significance of each of the four combinations of match and orientation. The results of these tests are shown in Table 14 below.

| Comparison | <i>t</i> | <i>p</i> |
|--|----------|----------|
| Mismatched images only: upright vs inverted | 14.172 | .000 * |
| Matched images only: upright vs inverted | 8.330 | .000 * |
| Upright presentation only: mismatched vs matched images | -1.894 | .058 |
| Inverted presentation only: mismatched vs matched images | -0.529 | .597 |

* = Significant at $p < 0.0125$

Table 14: Results of post-hoc t-tests on mean sadness ratings

p values were assessed at <0.0125 , and it was found that both mismatched and matched images evoked stronger feelings of sadness when presented upright compared

to inverted, but there was no significant difference in sadness between mismatch and match when only upright presentations or only inverted presentations were considered.

The effects of face mismatch and orientation on feeling disgusted

A 2*2 within-participants ANOVA was used to examine the two main effects of match versus mismatch and orientation on participant's reported experience of feeling disgusted. The main effects of match and orientation were both significant: $F(1,1077) = 551.843, p < .001$, partial $\eta^2 = .339$ for the match-mismatch factor and $F(1,1077) = 293.310, p < .001$, partial $\eta^2 = .214$ for orientation. There was also a significant interaction between match and orientation: $F(1,1077) = 298.051, p < .001$, partial $\eta^2 = .217$. The match factor had the largest effect size of the three, with the main effect of orientation and the interaction between the two factors having medium effect sizes. The interaction between match and orientation is shown in Figure 53 below:

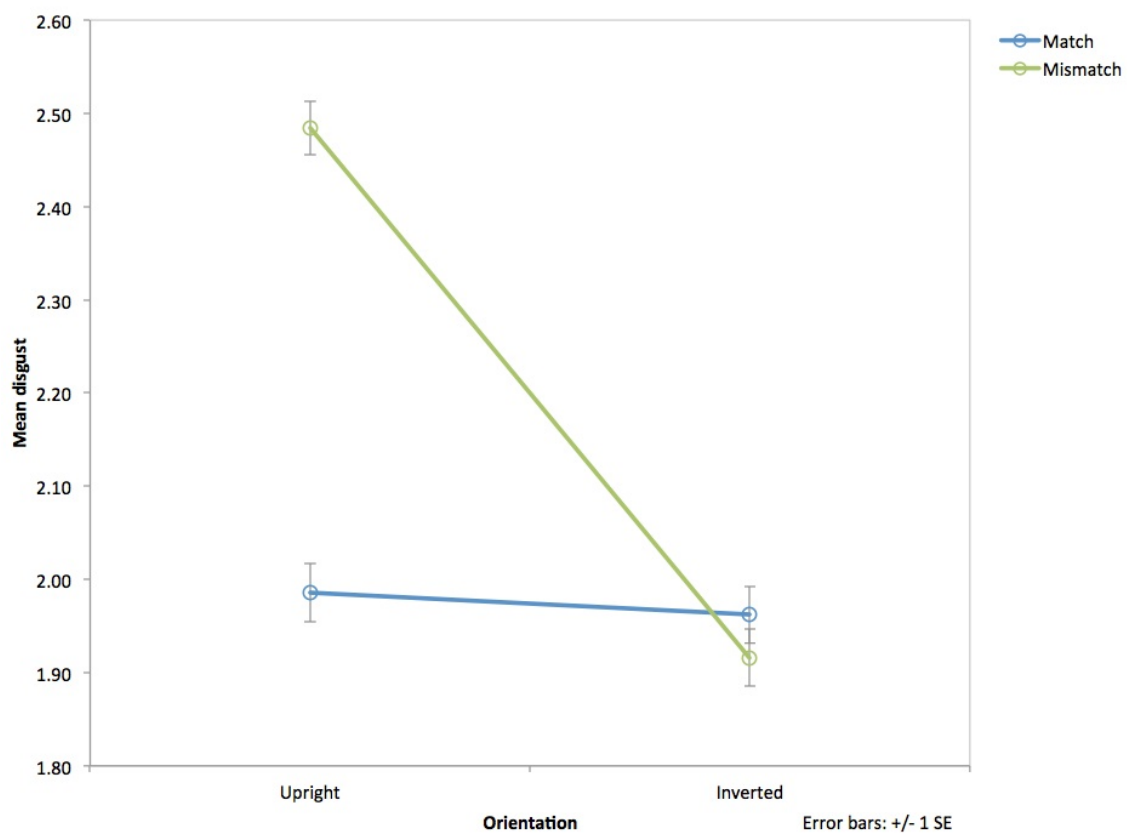


Figure 53: Comparison of the mean disgust strength for matched and mismatched face images when presented upright and inverted.

Post-hoc t-tests were used to explore the nature of this interaction by testing the significance of each of the four combinations of match and orientation. The results of these tests are shown in Table 15 below:

| Comparison | <i>t</i> | <i>p</i> |
|---|----------|----------|
| Mismatched images only: upright vs inverted | 9.917 | .000 * |
| Matched images only: upright vs inverted | 3.097 | .002 * |
| Upright presentation only: mismatched vs matched | 4.677 | .000 * |
| Inverted presentation only: mismatched vs matched | -0.753 | .451 |

* = Significant at $p < 0.0125$

Table 15: Results of post-hoc t-tests on mean disgust ratings

p values were assessed at <0.0125 , and it was found that disgust strength for both mismatched and matched images was significantly higher when presented upright compared to inverted. Mismatched images also evoked significantly stronger ratings of disgust than matched images when presented upright. However, match versus mismatch did not significantly affect disgust strength when images were presented inverted.

The effects of face mismatch and orientation on feeling happy

A 2*2 within-participants ANOVA was used to examine the two main effects of match versus mismatch and orientation on participant's reported experience of feeling happy. The main effects of match and orientation were both significant: $F(1,1077) = 343.934$, $p < .001$, partial $\eta^2 = .242$ for the match-mismatch factor and $F(1,1077) = 69.256$, $p < .001$, partial $\eta^2 = .060$ for orientation. There was also a significant interaction between match and orientation: $F(1,1077) = 111.152$, $p < .001$, partial $\eta^2 = .094$. While the differences between match and mismatch and orientation were significant, the effect sizes in this analysis were all small to medium, and the effect sizes for the orientation and interaction were very small. It was found that the match versus mismatch factor was responsible for

most of the variance in happiness. The interaction between match and orientation is shown in Figure 54 below:

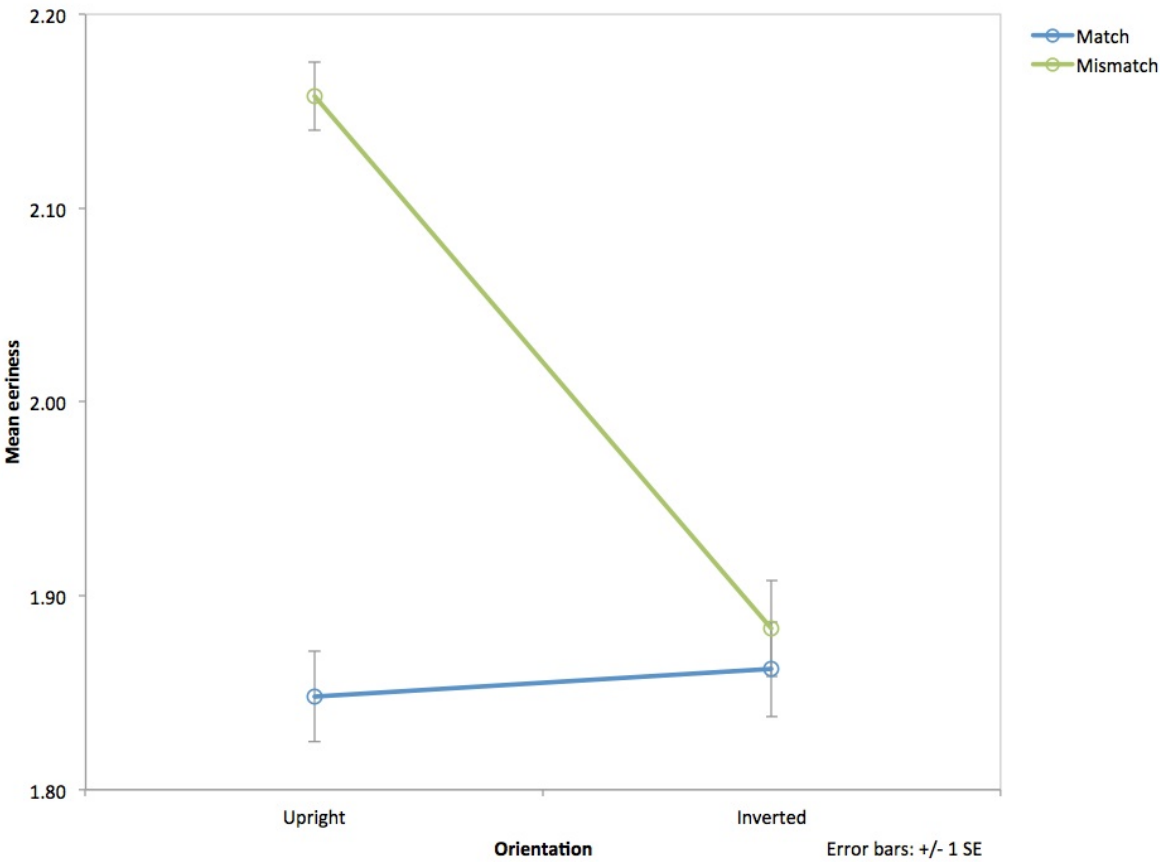


Figure 54: Comparison of the mean happiness strength for matched and mismatched face images when presented upright and inverted.

Post-hoc t-tests were used to explore the nature of this interaction by testing the significance of each of the four combinations of match and orientation. The results of these tests are shown in Table 16 below:

| Comparison | <i>t</i> | <i>p</i> |
|---|----------|----------|
| Mismatched images only: upright vs inverted | -7.719 | .000 * |
| Matched images only: upright vs inverted | -0.491 | .623 |
| Upright presentation only: mismatched vs matched | -2.891 | .004 * |
| Inverted presentation only: mismatched vs matched | 1.077 | .282 |

* = Significant at $p < 0.0125$

Table 16: Results of post-hoc t-tests on mean happiness ratings

Mismatched faces did evoke stronger feelings of happiness than matched faces, both when presented upright and inverted. Inversion was found to have a significant effect on happiness for mismatched images and mismatched images evoked significantly stronger ratings of happiness than matched images when presented upright.

H4, based on Bassili's (1979) work, predicted that certain mismatched blends would evoke strong ratings of happiness, disgust, anger, fear, sadness and surprise. The blends with happy or disgusted eyes should have evoked the highest ratings of those emotions, but both of those expressions seemed to be carried in the rest of the face, particularly in the mouth region as overall, the highest ratings of happiness were those given for the happy face, but highest blends were all of those with happy faces regardless of the eyes. Similarly, the disgusted faces with happy or neutral eyes both evoked slightly stronger ratings of disgust compared to the basic unmodified disgusted face.

The blends with angry or fearful faces should have evoked the strongest ratings of anger and fear. In terms of anger, the blend with the neutral face and angry eyes was rated as evoking the strongest response, with the same pattern found for ratings of fear with the neutral face and fearful eyes evoking the strongest fear rating within those blended faces.

Correlations between eeriness and emotional response.

Finally, the ratings of eeriness given for the eeriest blend (*happy face, fearful eyes*) were analysed to examine the correlations between this measure and each of the six emotions, exploring H5 and testing the links between uncanniness, fear and disgust. The eeriness ratings given for each participant when viewing the eeriest face were correlated with each of the six emotions. Significant correlations were found between eeriness and all emotions except happiness, visualised in Figure 55 below with the details of the correlation coefficients given in Table 17.

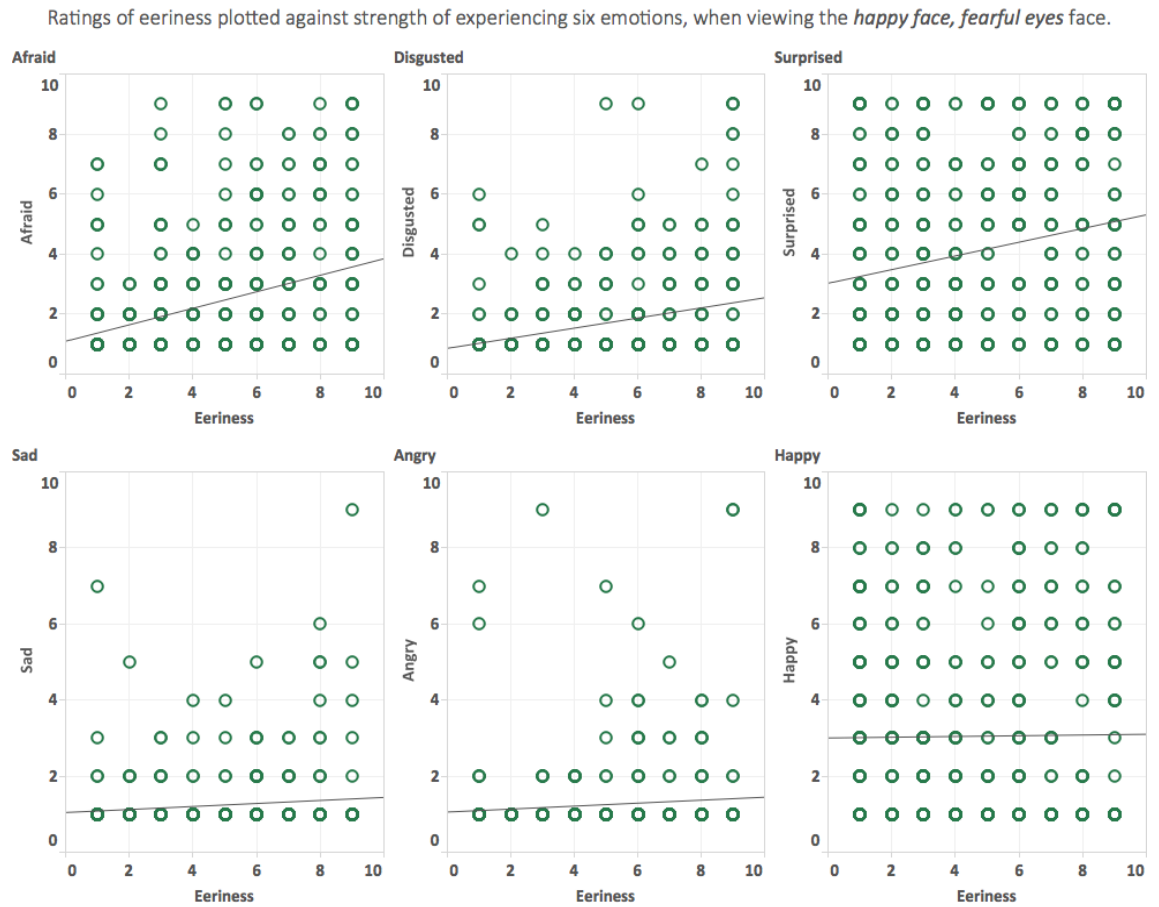


Figure 55: Eeriness ratings plotted against mean fear, sadness, disgust, anger, happiness and surprise.

| | <i>r</i> | <i>r-squared</i> | <i>p</i> |
|-----------|----------|------------------|----------|
| Afraid | 0.360 | 0.130 | < 0.005 |
| Disgusted | 0.327 | 0.107 | < 0.005 |
| Surprised | 0.217 | 0.047 | < 0.005 |
| Sadness | 0.138 | 0.019 | < 0.005 |
| Angry | 0.115 | 0.013 | < 0.005 |
| Happy | 0.010 | 0.000 | = 0.41 |

Table 17: Summary of Pearson's correlation tests on mean ratings of eeriness and fear, sadness, disgust, anger, happiness and surprise.

A positive relationship between eeriness and the negative emotions was found when participants were looking at the eeriest face, but no such relationship was found between eeriness and felt happiness. However, these were only very weak correlations and their

significance may have been an artefact of the large number of participants in this study.

These findings will now be briefly considered.

5.3.2.2: Discussion

The implications of the results of the analyses of emotional response ratings will now be briefly discussed to highlight key issues relating to the three hypotheses posed for this section.

Firstly, H3 proposed that there would be a difference between the matched and mismatched images on ratings of sadness, fear and disgust, and that this would mirror the finding reported in Section 5.3.1.1 above, where the eeriness ratings given to the mismatched faces were found to be eerier than the matched blends. Experienced happiness was also explored to see if this would reflect an inverted pattern compared to the eeriness ratings. These expected patterns were not found, firstly because the matched faces evoked stronger overall emotion ratings than the mismatched faces, but also as there was a significant interaction in experienced emotion between the match factor and the orientation in which the images were presented, and no such relationship was found for the eeriness ratings. Different patterns of interaction were found for the different emotions, suggesting that there is more complex relationship between orientation and mismatch for these emotions compared to eeriness. The effect of mismatch was to significantly increase eeriness in both orientations, and as fear serves as a useful benchmark to compare to the ratings of eeriness reported above it is interesting that the mismatched images were only found to be more frightening when presented upright, when they had been found to be eerier in both orientations. Sadness was felt significantly more strongly when presented upright compared to inverted for matched and mismatched images, and significantly higher ratings of disgust were given for mismatched

faces presented upright. These findings suggest that experienced eeriness does seem to be quite different from experienced fear, sadness and disgust. Happiness was also analysed to compare a positive emotion to these four negative ones, to see if an opposite pattern might be found, but this was not the case as it was found that inversion only had a significant effect on happiness ratings for mismatched images.

The ratings of emotion did not follow the patterns that were expected from the findings of eeriness but it should be noted that the differences between the reported emotion strengths were only very small, and while significant, their effect sizes indicated that this significance may have been an artefact of the very large number of responses received to the study. The mean emotion strength ratings for the matched and mismatched face in each orientation were also only rather low, with none of these means exceeding 2. Considering the rating scale used by participants, point 1 was labelled as 'I do not feel this at all' and point 5 was labelled as 'I feel this moderately strongly', so it cannot be argued that these mean ratings represent particularly strong experienced emotion for the matched or mismatched blends in either orientation.

H4, grounded in Bassili's (1979) work suggested that happiness and disgust will be experienced most strongly in the eyes of the blend, anger and fear in the surrounding face, but that there will be no distinct pattern for surprise or sadness. This was not supported from the results of this study as the emotions were most strongly evoked by the base images for those expressions, rather than in the blends that had been created by blending the eyes of those expressions into happy or neutral faces. The blends evoking the highest mean ratings of fear, sadness, happiness and anger were the un-blended base faces for those emotions. There was some overlap with another blend for the ratings of disgust as the *disgusted face, happy eyes* blend had the highest mean rating of disgust (4.4) with the unblended disgusted face image only slightly lower with a mean disgust of

4.3. Surprise was the only emotion reported by participants which was not actually present in the expressions used to create a blend, and this was included as some of the neutral and happy faces blended with the fearful eyes created a wide-eyed expression which was suggestive of surprise even though it wasn't originally present in either the base face or eyes. When analysed, it was found that the blends evoking the highest ratings of surprise were indeed the *happy face, fearful eyes* and *neutral face, fearful eyes* blends.

These mean ratings for each of the emotions evoked were also considered by looking at the emotions evoked for the blended images which had been found to be the eeriest in Section 5.3.1.1. The table below presents the mean ratings for the experienced six emotions, for those two blends which had been highlighted as particularly eerie, the *happy face, angry eyes* and *happy face, fearful eyes* images.

| | |
|--------------------------|----------------|
| Angry | |
| | Upright |
| Happy face, angry eyes | 2.0 |
| Happy face, fearful eyes | 1.2 |
| Afraid | |
| | Upright |
| Happy face, angry eyes | 2.3 |
| Happy face, fearful eyes | 2.3 |
| Sad | |
| | Upright |
| Happy face, angry eyes | 1.4 |
| Happy face, fearful eyes | 1.2 |
| Surprised | |
| | Upright |
| Happy face, angry eyes | 2.8 |
| Happy face, fearful eyes | 4.0 |
| Happy | |
| | Upright |
| Happy face, angry eyes | 1.6 |
| Happy face, fearful eyes | 3.0 |
| Disgusted | |
| | Upright |
| Happy face, angry eyes | 2.4 |
| Happy face, fearful eyes | 1.6 |


Mean emotion strength
1.0  5.0

Table 18: Mean fear, anger, disgust, happiness and sadness for the two eeriest blends.

It can be seen from this table that most of the evoked emotions were relatively low on the scale, with only the *happy face, fearful eyes* blend evoking an emotion above the bottom third of the scale for the ratings of feeling surprised. It is notable that even these eeriest blends did not evoke particularly strong ratings of fear, sadness or disgust. This does not provide the expected support for Ho et al's (2008) findings that these emotions were core components of the UVE.

Finally, H5 considered the correlations between participants' ratings of eeriness and their ratings of anger, disgust, fear, sadness and happiness. The hypothesis for this question was that ratings of eeriness should correlate strongly and positively with participant ratings of feeling fear and disgust but not with sadness or anger. Fear and

disgust have been previously anecdotally identified as emotions implicit in the uncanny valley effect, and this is also supported by the work of (Ho et al., 2008) who found that these were the terms most closely correlated with 'eerie' and 'creepy' in their inventory of the characteristics of NHAs. These variables were found to be correlated when analysed for the eeriest *happy face, fearful eyes blend*, as significant positive correlations were found between eeriness and all of the emotions except for happiness, with the strongest correlations found between eeriness and fear and eeriness and disgust. However, these were moderately weak correlations, explaining 13% and 11% of the variance respectively. While the correlations for eeriness and surprise, eeriness and sadness, and eeriness and anger were significant, there were only very weak, explaining 5%, 2% and 1% of the variance respectively.

5.3.3: Emotion identification

5.3.3.1: Results

Emotion identification: is accuracy of emotional recognition linked to face blend?

Having explored the ratings of eeriness for each blend and looked at correlations between eeriness and experienced emotions, an exploration of the accuracy of emotion identification was carried out. These results were drawn from the third question asked during Task 1 of Study 5. Participants were asked to identify the expression present in the face image by selecting one response from a list of anger, fear, happiness, disgust, sadness, surprise and no emotion. (See Table 12 for the full question wording and response labels.)

The accuracy of emotion identification was calculated as the percentage of correctly identified images for the matched and mismatched blends, presented upright and inverted. A correct identification for the unmodified face images simply meant correctly

identifying happiness, neutrality, anger, disgust, sadness or fear. Correct identification for a blended face meant identifying either the emotion present in the eyes or the emotion in the surrounding face. Percentage accuracy was calculated by dividing the number of correct identifications by the number of matched and mismatched faces seen in each orientation. It was hypothesised that the matched face image would be more accurately identified than the mismatched blends. A 2*2 within-participants ANOVA was used to examine the two main effects of match versus mismatch and orientation. The main effects of match and orientation were both significant: $F(1,1077) = 503.135$, $p < .001$, partial $\eta^2 = .319$ for the match-mismatch factor and $F(1,1077) = 151.877$, $p < .001$, partial $\eta^2 = .124$ for orientation. There was also a significant interaction between match and orientation: $F(1,1077) = 346.773$, $p < .001$, partial $\eta^2 = .244$. The interaction between match and orientation is shown in Figure 56 below:

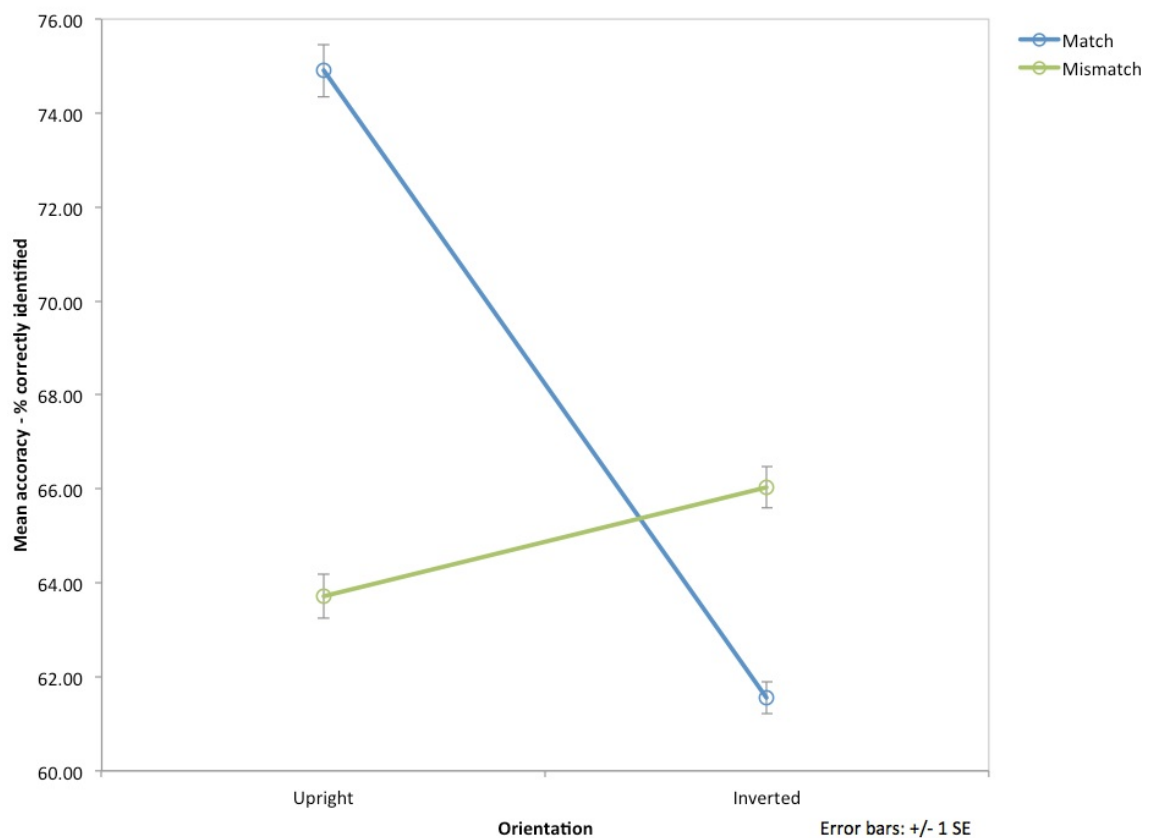


Figure 56: Comparison of the mean % accuracy for matched and mismatched face images when presented upright and inverted.

The matched faces were recognised more accurately than the mismatched faces when they were presented upright, and less accurately when they were presented inverted. However, the effect size for orientation was only small, with a medium effect size for the interaction and the largest was the main effect of match versus mismatch. To explore the interaction in more detail, post-hoc t-tests were carried out between all four combinations of match and orientation. The results are shown in Table 19 below:

| Comparison | <i>t</i> | <i>p</i> | |
|--|----------|----------|---|
| Mismatched images only: upright vs inverted | -6.746 | .000 | * |
| Matched images only: upright vs inverted | 11.011 | .000 | * |
| Upright presentation only: mismatched vs matched images | -29.471 | .000 | * |
| Inverted presentation only: mismatched vs matched images | -7.601 | .000 | * |

* = Significant at $p < 0.0125$

Table 19: Results of post-hoc t-tests on mean emotion identification accuracy

It can be seen that all comparisons were significant, suggesting that accuracy for matched images and mismatched images was significantly affected by inversion, and that when looking at the upright or inverted presentations only, match versus mismatch also significantly affected accuracy. Having found significant differences in accuracy between the matched and mismatched blends, the results were examined descriptively in more detail to consider the patterns of identification in each face image to explore H6 and H7. Figure 57 and Figure 58 below present heatmaps of the frequency of emotions identified, correctly and incorrectly, in each blend, when presented upright and inverted.

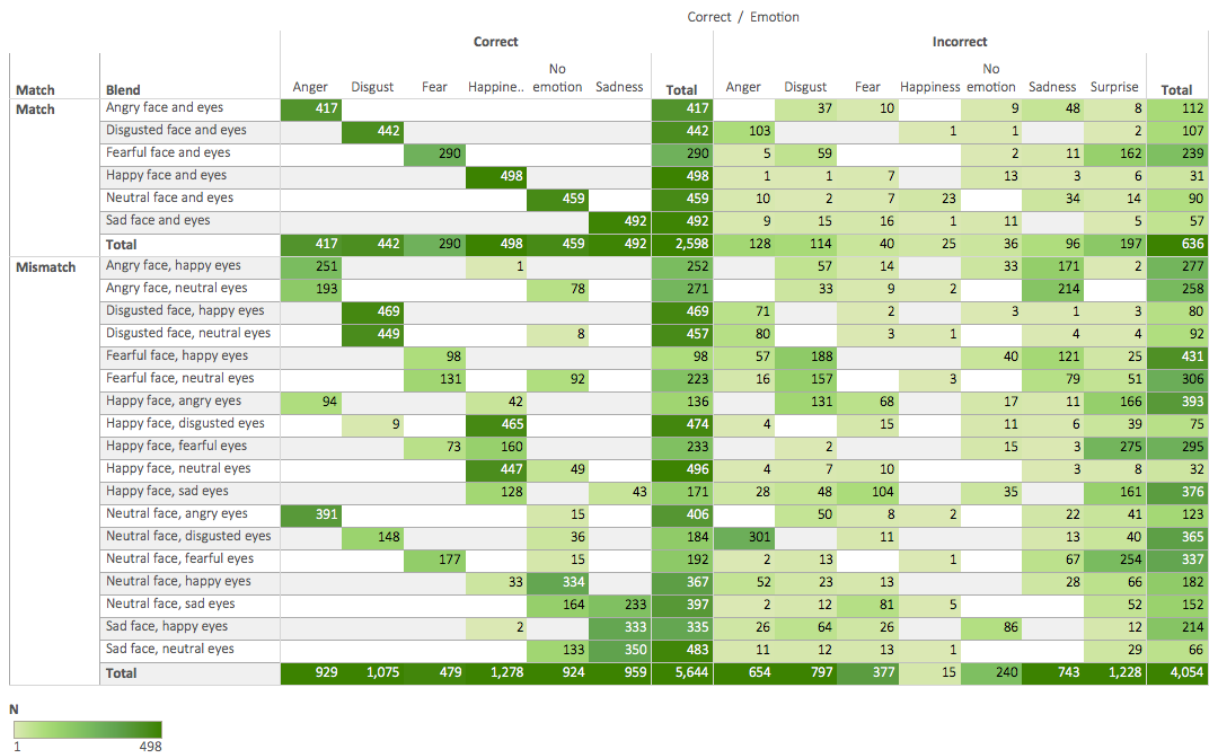


Figure 57: Heat map presenting the emotions identified for each blend, presented upright, indicating whether this was a correct or incorrect identification.

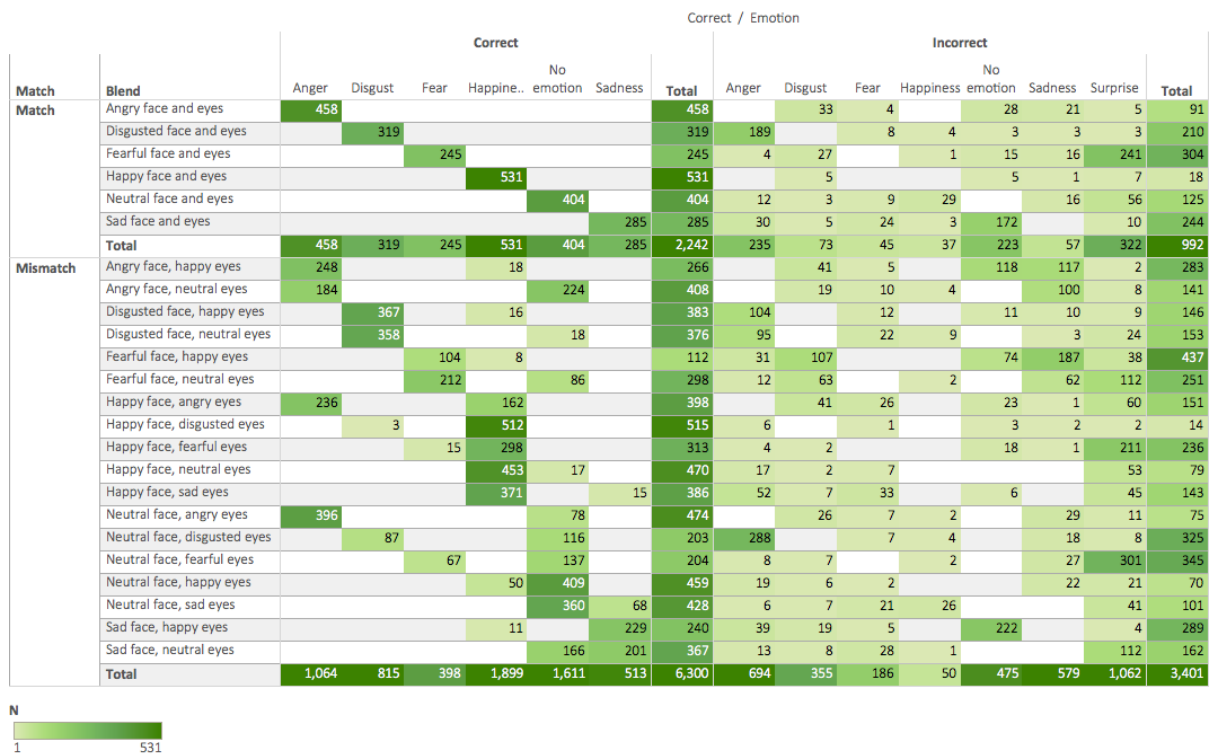


Figure 58: Heat map presenting the emotions identified for each blend, presented inverted, indicating whether this was a correct or incorrect identification.

These heat maps present the number of correct and incorrect identifications for each of the face images when presented upright and inverted. For example, it can be seen that there were 417 correct identifications of the original unblended face as angry when it was presented upright, compared to 112 incorrect identifications. The incorrect identifications also show the emotions that were mistakenly identified for each image, so again considering the upright angry blend, most of the incorrect identifications were that it was a sad expression rather than an angry one. The saturation of the cell background indicates the frequency identifications of that emotion, so by looking at the cells with the darkest and lightest green backgrounds it is possible to identify the key patterns of correct and incorrect identifications for each face image type. Upright faces containing happiness and sadness were the easiest to identify, with the highest frequency of correct identifications in their original unblended state and the highest correct identifications when blended with neutral eyes. Upright faces containing fear appeared to be the hardest to identify, with most incorrect identifications in their unblended state and also high numbers of incorrect identifications when fearful faces were blended with happy eyes. The unblended fearful face was most commonly identified as surprised, and the *fearful face, happy eyes* blend identified as either disgusted or sad. For the inverted faces, faces containing happy expressions were the most easily identified, with the highest frequency of correct identifications for the original unblended face but also high frequencies of correct identifications for the happy face with disgusted eyes. As with the upright presentations of the faces, fear was most commonly misidentified, both in its unblended state where it was most commonly confused with surprise and when the fearful face was blended with happy eyes, which was most commonly misidentified as sadness.

The heatmaps were used to consider the predictions made in H6, H7 and H8. H6 was based in Calvo and Lundqvist's (2008) findings and predicted that disgust and anger are

harder to distinguish and so disgust may be identified when a blended face contains anger and vice versa. Only partial support was found for this, as anger was not misidentified as disgust, but disgust was sometimes mistaken for anger. This occurred when images were presented upright, or when presented inverted. The highest frequency of incorrect identifications for the unblended disgusted faces was anger. Most of the misidentifications for the unblended angry faces reported sadness rather than disgust. Eight of the blended faces contained anger or disgust: *angry face, happy eyes; angry face, neutral eyes; happy face, angry eyes; neutral face, angry eyes; disgusted face, happy eyes; disgusted face, happy eyes; happy face, disgusted eyes; neutral face, disgusted eyes*. There were low frequencies of misidentifications of disgust for the anger-containing blends but the disgusted faces with happy or neutral eyes, and the neutral face with disgusted eyes, were both most commonly misidentified as angry. H7 was based in Calder et al's (2000) suggestion that recognition errors in different part of the face would vary by emotion type. It was predicted that the highest frequency of errors in identifying disgust and happiness would occur when these were present in the eyes of the blend, and that more errors in identifying anger, fear or sadness would be made when these were present in the face region. This was not supported in the results found in this study. Both when presented upright and presented inverted, the images with the highest error rates for disgust were those with the *fearful face, happy eyes* combination. Mistaken identifications of happiness were very low, with a maximum of five errors being made, which was for the neutral face with sad eyes. The upright blend most commonly misidentified as angry was the neutral face with disgusted eyes, for fear the happy face with sad eyes and for sadness the fearful face with the happy eyes.

H8 was a more general prediction, based on Hanawalt (1944) and predicted that happiness would be the most commonly identified emotion for the happy eyed blends,

and the negative emotions (anger, disgust, fear and sadness) identified most from the blends where those are the expression in the surrounding face. There were five blends where happy eyes were present: *angry face, happy eyes; disgusted face, happy eyes; fearful face, happy eyes; neutral face, happy eyes and sad face, happy eyes*. Of those, happiness was the most commonly identified emotion for only one blend, the *fearful face, happy eyes blend*. This was not found when the faces were presented inverted as fear was the most commonly identified emotion.

Emotion identification: are emotions more often correctly recognised from the eyes or the face?

To explore H9, this section will consider whether blended emotions will be identified correctly from the eyes or the surrounding face. The following heatmap is based on correct responses only and presents the frequencies of correct identifications by emotion, according to whether this was presented in the eyes or the surrounding face.

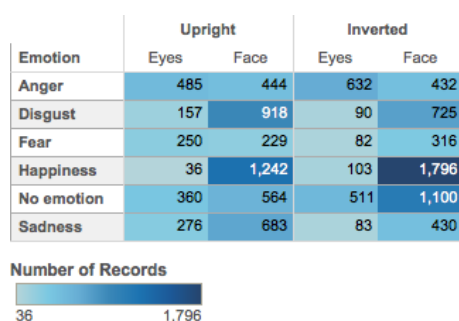


Figure 59: Comparison of correctly identified emotions by whether these were identified in the eyes or the surrounding face, when presented upright or inverted.

H9 was based on Bassili's (1979) work, and predicted that happiness and disgust would be identified in the faces where these are present in the eyes of the blend, and anger and fear identified where these are present in the surrounding face. There will be no distinct pattern of localisation for identification of surprise or sadness. This was only partially supported in the results of this study, as for both upright and inverted faces, happiness and disgust were identified from the face rather than the eyes. The prediction

for anger was supported for upright and inverted faces, as this were correctly identified more commonly from the eyes in both orientations, but not for fear where it was identified correctly from the eyes when upright and from the surrounding face when inverted.

A final notable finding was also considered, as this emerged from the analyses necessary to test H6-9. This was that the patterns of identifying happiness in the blended faces seemed to be particularly interesting for the two blends which were found to be eeriest and thus the best candidates for explaining the UVE. The results for the participants who had correctly identified any of the blends where a happy face was combined with negative eyes were extracted and analysed separately to compare the particular patterns of happiness identification for the two key eerie blends (*happy face, fearful eyes* and *happy face, angry eyes*) against the other three blends where a happy face had been blended with neutral, disgusted or sad eyes. A correct identification could be made either from the face or the surrounding eyes and the heatmap highlighted that in the case of the two eeriest blends, participants were less likely to make their correct identification based on happiness and more likely to base it on the angry or fearful components. The segmented frequency chart below compares these re-categorised variables:

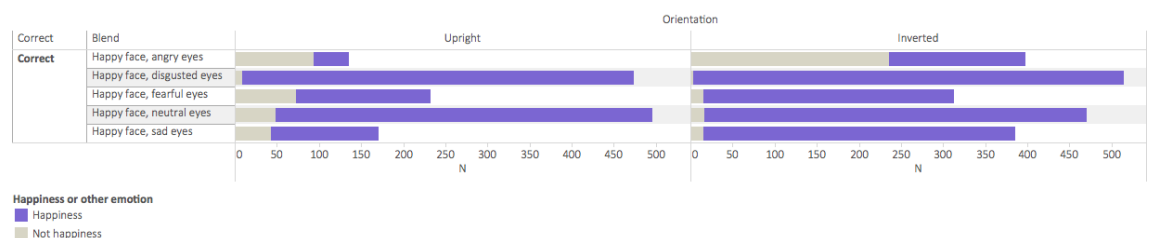


Figure 60: Comparison of happy/non happy emotions identified in blended faces where happy faces were combined with different types of non-happy eyes.

Happiness was the emotion recognised more commonly in all of the blends but the angry, neutral and sad blends did have a smaller identification from happiness than the

happy faces with disgusted or neutral eyes. The implications of these observations will be discussed below.

5.3.3.2: Discussion

Three areas of emotion identification were considered in this section. Firstly, emotion identification accuracy was analysed to see whether it would be affected by the type of face image (either matched or mismatched) and the orientation in which it was presented. It had been hypothesised that the matched images would be easier to recognise than mismatched images, and this was found, but more importantly an interaction effect was found so this relationship was only the case for images when presented upright. As a result, orientation was considered when examining each of the three follow-up hypotheses which explored the nature of this difference.

These three follow-up explorations of the nature of the accuracy of identifying emotions from the blended faces were carried out with reference to three previous theories of emotion perception. The results of this study did not provide full support for any of the theories that were proposed. Firstly, only partial support was found for Calvo and Lundqvist's (2008) predictions of complementary confusion between anger and disgust, as blends containing anger were not consistently misidentified as expressions of disgust, but blends containing disgust were misidentified as angry, both when presented upright and when presented inverted. Calder et al (2000) predicted specific patterns of emotion identification from different face regions but none of the blends in this study supported those patterns. Finally, support was also not found for Hanawalt's (1944) proposal that generally, happiness would be perceived in faces with happy eyes as only one of the blends (*fearful face, happy eyes*) had particularly high identification based in the eyes.

It may be that there was some aspect of these blended faces that caused participants to respond in a new and unique way, but these findings may also have been a result of applying previous theories based on the perception of unmodified faces to these blended faces as Calvo and Lundqvist and Hanawalt both used images of natural human expressions in their research. Only Calder's study used manipulated faces, and as these were top-bottom chimeric faces it could be argued that the results of this study would be most likely to follow those predictions.

A final exploration moved beyond the stated hypotheses to look at an unusual property of how emotions were identified from the blends where negative eyes were blended into the happy faces, as the two blends that were identified as particularly eerie in Section 5.3.1.1 were mostly identified from the negative eye region rather than the happy surrounding face, and this was not the case for the other negative eyed blends. It may be that the eeriness of these two blends occurred as a result of it being harder to perceive positive emotions from these faces, or that for some reason participants chose to focus particularly on the negative features of these faces rather than the positive, happy component. The data collected in this study do not support a further exploration as to whether these explanations may be true but it does suggest an interesting area for future research.

5.3.4: Emotion classification

5.3.4.1: Results

Emotion classification: is the ability to categorise emotional expressions linked to face blend?

A final analysis briefly considered how participants would classify the blended emotions, to explore which of the blended emotions would be used in the categorisation,

and whether the eerier faces would be harder to categorise than the other blends. These results were drawn from Task 2 of Study 5. Participants were presented with twelve blended faces and asked to drag and drop each image into one of four boxes on screen to indicate whether they felt the emotion it presented was broadly positive, negative or neutral, with the final box used for those emotions that were too hard to classify. Participants were also asked to rank the images within the four categories in order of how well they represented positive, negative or neutral expressions. However, some of the categories were used for very small numbers of images so the ranking data was not used for further analysis. The images in Task 2 were all blended faces, not the original unmatched images, and they were all presented upright. The images in Task 2 were created from different models from those used in Task 1 so although participants had seen those types of blended face in the first study, they had not seen those specific face images. (See Table 12 for the full question wording and response labels.)

A segmented frequency chart was constructed to visualise the distribution of categories used for each of the blended faces:

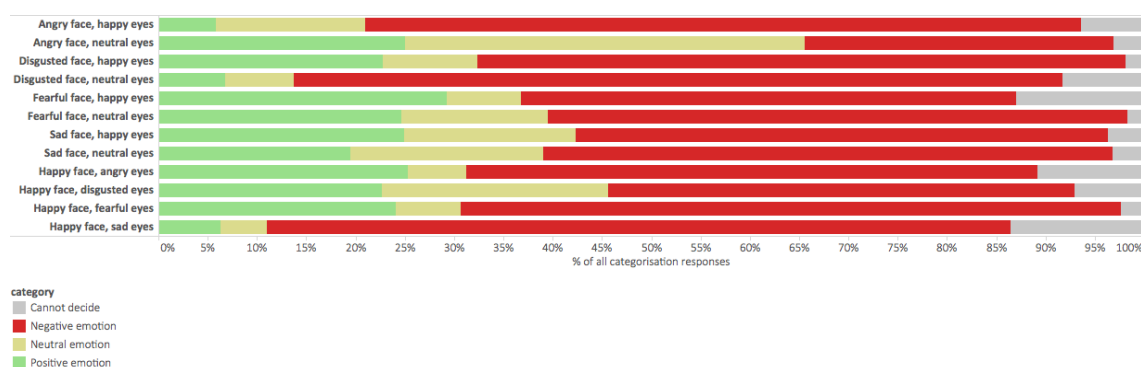


Figure 61: Categorisation ratings for the blends presented.

It was hypothesised that the eerier blends would be the hardest to categorise. The table below shows the proportions of the ‘cannot decide’ category compared to the other three and contrary to expectations, the *happy face, sad eyes* blend was found to be the

hardest to categorise, measured by the percentage of participants who indicated that they were unable to assign a blend to a category.

| | All other categories | Cannot decide |
|------------------------------|----------------------|---------------|
| Angry face, happy eyes | 93.6% | 6.4% |
| Angry face, neutral eyes | 96.9% | 3.1% |
| Disgusted face, happy eyes | 98.1% | 1.9% |
| Disgusted face, neutral eyes | 91.7% | 8.3% |
| Fearful face, happy eyes | 87.0% | 13.0% |
| Fearful face, neutral eyes | 98.3% | 1.7% |
| Sad face, happy eyes | 96.3% | 3.7% |
| Sad face, neutral eyes | 96.8% | 3.2% |
| Happy face, angry eyes | 89.2% | 10.8% |
| Happy face, disgusted eyes | 92.9% | 7.1% |
| Happy face, fearful eyes | 97.7% | 2.3% |
| Happy face, sad eyes | 86.4% | 13.6% |

Figure 62: Comparing 'cannot decide' and other categories by blend.

The implications of these findings will now be briefly discussed.

5.3.4.2: Discussion

The classification task in this study has been included to see if there are any patterns in participant ability to classify blend faces in terms of the emotions they display. Three hypotheses were initially proposed. Firstly, as noted above, Hanawalt (1944) suggests that positive emotions are generally perceived from the bottom of the face, negative from the top. The first hypothesis is that each of four different versions of the *happy face, negative eyes* blends would be classified as displaying positive emotions but this was not supported from the classifications carried out in this study where participants all classified those as displaying negative emotions. This was particularly strongly shown for the *happy face, fearful eyes* and *happy face, angry eyes* blends which have also been identified as the most eerie of all the blended emotions.

Going on to test the second part of that hypothesis, that the *negative face, happy eyes* blends will be categorised as displaying negative emotions found that all four blends

(*angry face, happy eyes, disgusted face, happy eyes, fearful face, happy eyes and the sad face, happy eyes*) were indeed classified as negative.

A final question to consider was that of whether the blends where neutral eyes were combined with any of the other base faces would be particularly hard to classify. This was a development from Tinwell et al's (2013) hypothesis that these would be the eeriest blends and also those presenting the most ambiguous appearance to participants. It is hypothesised that these unsettling faces would be the hardest to classify as their unusual nature may make it harder for participants to make decisions about the category to which they belong. Most participants were able to classify the blends, with only a small number using the 'cannot decide' category for each blend. However, when the percentages of participants who said they were unable to decide about the face category were compared to those who were able to classify the blends, the neutral eye blends did not emerge as those where it was hardest to decide as that blend was the happy face, sad eyes blend. It may be that there is a relationship between the extremity of mismatch and classification ease, as faces where happiness and sadness were combined would make for the most extreme pairing. However, this relationship is unlikely to contribute to an understanding of the UVE per se as neither the *happy face, sad eyes* nor the *sad face, happy eyes* blends were found to be particularly eerie when those ratings were gathered.

5.4: General Discussion of Phase 3

This final research phase took a core finding from the results from Phases 1 and 2 that certain types of NHF are eerier than others, and developed this by testing whether this could be explained by looking at how NHFs are able to convey emotional expressions. It explored the idea of emotion mismatch to test the hypothesis that expectations of affinity and empathy may be established in an overall impression of a NHA but when these

cannot be fully realised from the finer detail of the appearance, this may cause the uneasiness that characterises a UV response. An examination of the relevant face perception and UV literature suggested four research questions and associated hypotheses, and an online study was designed to collect respondent data to measure this. These sources had suggested a wide range of potential outcomes for different types of mismatch so a large set of novel stimuli were designed which blended together different emotional expression from eye and faces regions, allowing a detailed and complex explorations of the interactions between different emotions. The blended faces were presented upright and inverted to continue the explorations from Phase 2 and examine which components of emotion expression processing would be disrupted by the inversion of the image.

The key finding from this phase was the discovery that blending emotional expressions did indeed have a significant (yet subtle) effect on participant perceptions of eeriness and how they felt, with clear patterns emerging that the blended faces were certainly more eerie than any of the unblended faces; a finding that persisted whether the face was inverted or not. However, the actual ratings of eeriness were low so it cannot be concluded that participants found any of these faces particularly eerie. Overall, the eeriest faces were those where the happy faces were combined with the negative emotions, particularly the two blends where fearful or angry eyes were blended into the happy face. Relating back to the question from Phase 2 on analytic versus holistic processing, it is notable that every one of the inverted faces were rated as eerier than the upright faces, whether blended or not blended. One explanation for this may be because the usual orientation for encountering a face is upright and so all faces when presented inverted look stranger than they otherwise would do, so even with the unblended faces where no unusual features were present in either orientation, an expectation of strangeness may be

driving the higher rating of eeriness. The eeriness ratings were supported by participant ratings of how strongly they felt different emotions, with the *happy face, fearful eyes* and *happy face, angry eyes* blends also rated as evoking the highest ratings for feeling afraid.

Finally, an analysis of the emotion identification data found that a feature of those two particularly eerie faces was that participants were focusing on the eye region when making their decision. All of the blended faces were created from two base images, but participants were not paying as much attention to the happy part of the face in those blends with angry and fearful eyes, compared to the other happy-face blends. It is hypothesised that an extreme mismatch between the expressions present in the eye and face regions of these blends seemed to over-ride the happiness that could have been perceived from part of the face displaying a happy expression. These eeriness ratings for the happy faces with fearful or angry eyes blended into the image may be an indication that participants found those blended faces to be more aversive because they did not feel such a sense of affinity towards them.

Consideration will now be given to aspects of the experimental design which may have had a confounding effect on the results. Firstly, the nature of the stimulus images themselves should be considered, especially as it was observed in Phase 2 that the stimulus images were clearly the product of a computer transformation and would be hard to encounter in real life. The blended images used in this third phase were developed with this in mind, and care was taken to try and blend the images carefully so they were not immediately obvious as manipulated images. In this way, it should be possible to generalise those findings to real world encounters with NHAs displaying mismatched expressions, as well as those presented as static images. These Phase 3 images do avoid the issues raised for those used in Phase 2 faces as at first glance, many

of the faces did not appear that unusual and it was sometimes hard to tell just which expressions had been used to create each blend.

It can be concluded that this research phase has some indications that started to provide some evidence suggesting that the sense of eeriness that characterises the UVE may be caused by certain types of mismatched expressions. In this study, faces which have happy, smiling expressions in the lower part of the face along with expressions in the eye region which are suggestive of anger or fear were rated as the most eerie, but it may represent a more general effect that eeriness can be elicited where there is a marked disconnect between the expression that is shown in the face and the eyes. These findings are framed as initial evidence that this may be a generally applicable effect but would need to be tested further by looking at different participant cohorts and other types of face processing task to evaluate the extent to which it can be considered as a generally applicable finding. In terms of applying these indications to a real-world context, it can be concluded that designers should attempt to avoid ambiguity of facial expression between different parts of a face, by matching levels of detail and realism between eye and face regions. For example, concentrating on highly detailed smiles in a face where the eyes could not convey the same level of happiness may cause people to reject the agent, but considering the types of mismatch found to be particularly eerie, this could be made even worse in animated agents where expressions are changing and may not always be well synchronised. A fearful or angry expression that was still being displayed in the eye region while the mouth had moved on to presenting a smile may be unsettling and so should be avoided.

The General Discussion section to follow in Chapter 6 will consider the UVE in general, and suggest areas where research in this area could be progressed, but several fruitful areas have been suggested following up on the findings from this phase alone. Empirical

testing of the expression 'lag' suggested above is needed to explore whether the extent to which what has been observed in the image blends could be replicated in an animated agent. It would also be useful to examine participant sensitivity to the different blends, and look for what would constitute a just noticeable difference in eeriness as different expressions are blended, and whether it might be possible to map the point at which eeriness emerges when different features are blended together. Finally, repeating this study but taking a cross-cultural perspective was initially proposed by Brenton et al (2005) who suggested that the UVE is culturally dynamic. Brenton et al gave the example of facelifts, which generally do not cause consternation in this culture but may in one where that sort of surgery is not commonplace. It can be imagined that the discrepancy in perceived age of features between a young face and aged hands might be unsettling if you were unaware of the reason why. Similarly, it could be argued that there may be cultural differences in the extent to which mismatched expressions are acceptable, as to an extent some of the expectations that it is believed are being challenged here, leading to rejection of the mismatched faces, are set up are based on what you expect to be able to read from a face. For example, some cultures are attuned to more subtle facial expressions than others, with the effect that it is hard from people outside that culture to read subtle expressions. Recent research by Yuki et al (2007) and Koda and Ruttkay (2014) looked at the region of the face that was most influential in identifying emotions from cartoon-like faces, and found significant differences between Hungarian and Japanese participants. They found that the Japanese participants consistently paid more attention to the eye region compared to the mouth region, with the opposite being found for the Hungarian participants. It would be very interesting to look at whether the same eeriness, emotional response, identification and classification patterns found in this study would also be found in non-Western participants. While participant demographics were collected for this

phase (see Appendix 3a for a full breakdown) there was insufficient diversity in the respondent population to break down the response data in a meaningful sense as the majority of the participants in this study (86%) identified as White, either British or Other. A modified version of this experiment, directed to participants from different cultural backgrounds, would be highly useful in exploring the extent to which these findings can be generalised and also to add to knowledge on the design of virtual agents.

Chapter 6: General Discussion

6.1: Overview of Research Findings

The overarching aim of this thesis was to explore the UVE , which began by examining the concepts described by Mori (1970, 2005, 2012), Pollick (2005) and MacDorman (2006) and using these to design a novel method for collecting evidence which would allow a mapping of the relationship between human-likeness, strangeness and eeriness alongside an exploration of the subjective experience of encountering NHAs. The findings of Phase 1 suggested that there were indeed interesting qualities in NHFs that warranted further exploration, and once the nature of those relationships between perceived human-likeness and emotional response had been established, later studies were designed to step off those familiar paths within the UV, and move into the as-yet unexplored territory of how NHFs were perceived. Several novel and notable observations can be made from these explorations, and this general discussion will consider these across five areas. In this overview of the research findings, the three key theories explored in this thesis will be revisited and the findings discussed in the context of what they are able to contribute to an understanding of those concepts. These will then be discussed with specific reference to approaches to researching the UVE, looking back across the various explanations that had been proposed for the UVE and discussing the extent to which the present thesis has addressed the vexing issue of circularity in this area. This leads into a third section where the methods used in this thesis will be critically considered. A fourth section on future directions for research in this area will be considered, and finally this thesis will close with some concluding thoughts.

The first theory to be reviewed revisits the original concept of an UVE based on Mori's 1970 idea that emotional responses to artificial agents would become more positive as

those agents became more human-like but only up to a certain point when they would start to appear unsettling rather than likeable. Notably, this point of rejection would be typified as uncanny and characterised by a sense of unease and feelings of eeriness. MacDorman's (2005c) interest in this area brought the concept to a new level of prominence and initiated research in various disciplines, but at the heart of the present research into the UVE has been the hypothesised dip in positive responses at the near-human point. Three of the studies in this thesis compared measurements of human-likeness with subjective ratings of emotional response and while a non-linear pattern was found in all of the analyses, these studies did not find an UVE where high eeriness accompanied levels of human-likeness which were almost but not quite human. In the first of two studies in Phase 1, the strangeness and eeriness ratings for 20 NHFs found that the eeriest faces were those that were not human, or had a moderately human appearance. This was confirmed in the second study in Phase 1, where three faces taken from the first study were compared to each other and also to human and artificial anchor faces. While a non-linear pattern was found, the eeriest and strangest face was the moderately, rather than highly, human face. This was also found in Phase 2, where rather than being drawn from a collection of human-like faces rated for their human-like appearance, an attempt had been made to systematically vary the human-likeness by morphing between different types of non-human face and matched human face images, and evaluating still images taken at different points.

The UV hypothesis would predict that the nearly-human point in the morph (where the face represented 75% human-likeness and only 25% 'otherness') would have been rated as the most eerie and most strange, but the eeriest face was found to be the one at the 50% point. While this was an unexpected finding, it is not sufficient to reject the UV hypothesis as the overall relationship between human-likeness and emotional response

did consistently follow the predicted pattern that positive responses would generally increase as human-likeness increased, but that there would be a point at which the positive ratings dipped rather than continuing to increase, and it is at this point that it was predicted the eeriest faces would be found. Instead, the eeriest faces were found to be those which were 50% artificial and 50% human. This is not the only study of the UVE to have identified the mid-point between artificial and human in their findings. Cheetham et al (2013) and Burleigh et al (2013) both identified this as the key location rather than the near-human point. This may also reflect Tinwell's (2009) and Tinwell et al's (2011a) findings which described a series of valleys within the overall UV plot. These appeared as several dips in familiarity at different points of human-likeness and were clustered around different types of NHA. It should also be considered that the finding may be as a result of the selection of individual stimuli for rating, or the relatively rough-grained number of morph stages. In the latter case, introducing more steps around the moderate to highly human-like morphs may allow finer discrimination in human-likeness.

The second part of Mori's (1970) hypothesis was that the emotional response will be characterised as eerie and the present research was able to explore this in each of the phases by measuring ratings of eeriness. Mean ratings of eeriness varied across the phases, with the eeriest ($M=9.6$) face found to be the dark eyed doll-face chosen as the Cluster 1 exemplar in Phase 1 and the least eerie ($M=1.7$) the unmodified base sad face used in Phase 3. Figure 63 below illustrates this range by presenting the images in the top and bottom 10% of mean eeriness. Table 32 in Appendix 4 presents the full list of all stimulus images and their mean eeriness.

| Phase | Face description | Orientation | Eeriness rating |
|-------|---|-------------|-----------------|
| 1 | Cluster 1 Face (Low human, highly strange) | Upright | 9.6 |
| 2 | Animal - 50% human | Upright | 7.7 |
| 1 | Cluster 2 Face (Medium humanlikeness and strangeness) | Upright | 7.0 |
| 2 | Robot - 50% human | Upright | 6.8 |
| 2 | Robot - 75% human | Upright | 6.4 |
| 2 | Robot - 25% human | Upright | 6.3 |
| 2 | Animal 75% human | Upright | 6.1 |
| 3 | Happy face and eyes | Upright | 2.1 |
| 3 | Happy face, disgusted eyes | Upright | 2.0 |
| 3 | Neutral face, angry eyes | Upright | 2.0 |
| 3 | Sad face, neutral eyes | Upright | 2.0 |
| 3 | Neutral face, sad eyes | Upright | 2.0 |
| 3 | Neutral face and eyes | Upright | 1.9 |
| 3 | Sad face and eyes | Upright | 1.7 |

Figure 63: Eeriness ratings of face images presented in each research phase, indicating those rated as most and least eerie.

It can be seen that some of the images, those highlighted in green as the top 10%, were indeed rated as eerie. However, 46% of the images had a mean eeriness of 3 or below, and so although differences in eeriness were found, not all of the stimulus images were rated as strongly or highly eerie, so some of the analyses in this thesis have been based on relatively small differences in eeriness. This wide variation in eeriness is an important point to consider, given the crucial role that eeriness plays in any discussion of the UVE. While it has been acknowledged above that eeriness is a subtle and complex emotion, and so reported experiences of eeriness could be expected to be lower than those for outright fear or startle responses to unexpected stimuli. However, the range of responses across the different phases is interesting and suggests that further research in this area may be able to shed light on why this may be. A working theory is that the experimental settings for all of these studies were designed to be non-threatening, and by presenting static images of faces to participants it meant their comfort levels made it difficult for them to experience being genuinely unsettled by the images that they were viewing. The external context for the research is also worth considering at this point, given that ratings of eeriness decreased across the different phases and these studies

were conducted over a period of several years. It may have been that the general population simply became more acclimatised to the experience of looking at artificial faces due to greater exposure in day-to-day life, setting general expectations for eeriness at different levels over time.

It can be seen that a valley effect was found in the relationship between human-likeness and emotion response, and even though the topography of this valley was not as expected, some of the stimuli were able to elicit uncanny responses. Later studies in this thesis looked in detail at mechanisms that may have been responsible for this response pattern. It can be seen from the scope of the literature presented in the Chapter 1 that there were a range of theoretical perspectives and approaches that could have been adopted to explore the UVE. This thesis began with a broad exploration of the effect, which was then refined to test two specific areas of face perception theory and research which were applied to the UVE. The first of these applied face perception theories to examine how information about the appearance and identity of NHFs was encoded, using a recognition task to look at whether performance would differ as a function of human-likeness.

Previous studies have found that one cause of the UVE was the presence of anomalous, unrealistic or otherwise unusual features, most notably eyes. Seyama and Nagayama (2007, 2009) found that manipulations to the size and realism of eyes could explain uncanniness in a range of different face types. Distorted or manipulated features have been used in face perception research to isolate and explain mechanisms of face processing and information encoding, with Searcy and Bartlett (1996) use of grotesque features and Thompson (1980) and Cornes et al's (2011) technique of inverting features within a face have both having been used as evidence that upright face information is encoded at the level of the whole face rather than individual features, and that faces are

processed holistically rather than analytically. Natural human faces are known to be processed differently from other types of complex object (Kanwisher and Moscovitch (2000)) but no research had been carried out into how this would apply to NHFs. The idea of a transition between 'non-human' and 'human' is key to the UVE but the concept that at some point on that continuum there may also be a switch in processing mechanisms where the increasingly human-like faces stop being processed as objects and begin to be processed as natural faces had not been explored. The presence of a FIE (Yin, 1969) at different levels of human-likeness for different categories of face type was used to examine this, and to specifically test whether some of the NHFs (with the most unusual features) would show a decreased inversion effect suggestive of analytic rather than holistic processing. Hills et al's (2011) 'feature salience' findings also suggested that the morphed faces would not be as affected by inversion as the natural human faces, as in their studies, selectively directing attention to the eye region meant that there was not as much of a performance decrement when those faces were inverted. While Study 4 did not use a cueing technique to direct attention in that way, it may be that the unusual appearance of the morphed eyes could have served a similar purpose in directing attention to that area. The results found that an inversion effect did occur at all levels of human-likeness, but the predicted effect of human-likeness on that inversion effect was not found. It had been expected that the morphed faces would not be as affected by inversion as the natural human faces, and so they would have been recognised more quickly, but the *slowest* performance was actually found for the faces at the 75% human point where they were almost but not quite human. The suggestion that the anomalous features may have 'protected' these NHFs from the detrimental effects of inversion could not be supported. One explanation for this may be in the nature of the changes in the appearance of the faces that occurred when the faces were morphed from non-human to

human, particularly considering these findings in the light of Searcy and Bartlett's (1996) observations of the impact on inversion of manipulations that change faces featurally or configurally. They found that the inversion effect was least for the faces which had been disrupted in their configuration, and most for the ones where the features had been altered. The larger inversion effect found for the morphed faces may be explained in this study as morphing keeps the configuration of the face static while incrementally altering the appearance of the individual features which does mirror the techniques used in Searcy and Bartlett's (1996) study. Cheetham et al's (2013) finding that participants spent longer looking at objects at the perceptual boundary between artificial and human may also explain why participants spent longer looking at those morphed faces as they would certainly be ones that straddle the boundary between the human and non-human.

Additionally, some explanations for the increased inversion effect for the morphed faces may be found in studies which offered interpretations of the FIE that did not rely exclusively on analytic versus holistic processing. Rhodes et al (1993) found a difference in the size of the inversion decrement depending on which parts of a face had been manipulated and how they were presented. Their results showed that the largest decrements occurred when individual eye and mouth areas were isolated and presented without the context of the surrounding face. While this study used whole faces rather than presenting features in isolation, the markedly unusual appearance of the eye regions in the morph images may have drawn attention selectively to those areas, resulting in the large inversion decrement that occurred for the morphed faces. Sekuler et al's (2004) theory suggested that the ease of extracting information from a face was key to whether an inversion effect occurred and that it was not the mode of processing which was relevant, but the ability to extract the information from the faces when they are inverted. It could be that the morphed features of the NHFs represented a particular challenge for

information extraction, meaning that it took longer to discriminate between them. Finally, it may be that the non-human starting points for the face morphs could not really be considered as objects rather than faces. While the robots and statues could perhaps be read as object-like before the transition to human-likeness, the dolls and animals are clearly identifiable as faces, albeit non-human ones. An area for further exploration would be to repeat this study and use some of the most abstract face-like representations currently in use for some embodied agents as the non-human starting point to see if this has an impact on the way faces drawn from that continuum are processed.

The second theoretical construct applied to the question of the UVE in this thesis was the question of whether of mismatched and incongruent emotional expressions would elicit perceptions of eeriness. This construct drew upon three sources of theory and evidence in formulating the research questions and the experimental design. These were the literature relating to the UVE as an artefact of categorical perception, empirical studies of cue mismatch and the impact of incongruent stimuli on the UVE, and face perception literature looking at how emotional expressions are perceived. It was also a development of the strand mentioned above where the eyes had been identified as a probable cause of eeriness to ask whether the reason that large or distorted eyes make people feel uneasy may be because they interfere with an ability to present a realistic, relatable emotional expression. The general prediction was that the eeriest faces would be those with a mismatch between the expression shown in the area immediately around the eyes and that suggested in the rest of the face (e.g. the expression suggested by the mouth position, shape and openness or by creasing of the forehead). In terms of categorical perception, this follows Ramey's (2005) prediction that uncanniness emerges when entities straddle categorical boundaries because they possess attributes from two possible categories, making it hard to accurately allocate them to either category. The

presence of two different expressions in one face may make it hard to categorise the emotion shown by the person, causing a difficulty in classifying them as, for example, definitely happy or definitely angry.

Cue mismatch theory is recognised as an area of enquiry in the UVE, with research by Seyama and Nagayama (2009) and MacDorman et al (2009) having found that mixing levels of detail in a virtual agent elicited eeriness and negative evaluations as the agent was not able to present a realistic overall appearance. MacDorman (2009) also found that the more extreme this mismatch, the more negative the overall assessment. Tinwell's research into cue mismatch has been extensive and has found evidence for UVEs in animations where the spoken dialogue was not perfectly synchronised with the moving image (Tinwell, 2009), in videos of faces with a mismatch between the emotional expression conveyed in the voice and that suggested by the expression (Tinwell et al, 2010), and a lack of expression in the upper part of the face (Tinwell et al, 2011b, 2013). The latter explanation was key to the design of Phase 3, as Tinwell et al had found that faces lacking emotion in the upper part of the face while displaying emotion in the lower part of the face were the eeriest, but one aspect that had not been explored in Tinwell's studies was the scenario where different expressions were presented in the two parts. Tinwell et al's explanation for the eeriness of a lack of upper-face emotion drew upon the perception of psychopathic traits from facial expressions, where expressions are only conveyed in the lower part of the face and it was suggested that these faces may seem eerie because they elicit the same warning signals. This idea was developed in the present thesis by considering Ekman et al's work on emotional expressions (particularly Ekman et al, 2005) and how they are perceived, predicting that an extreme mismatch where completely incongruent expressions were presented in the same face may augment this understanding even further.

Blended faces were created from images of two different emotions so the eye region always displayed an expression incongruent with the rest of the face. These were used (along with their unblended originals) in a series of tasks with participants where they were asked to rate them for eeriness, indicate how strongly they elicited different emotions and identify the emotions that were present in the face. To continue the FIE strand in Phase 2, faces were presented inverted as well as upright. The results supported the prediction that mismatched faces are eerier overall than matched ones, and it was also found that they were still eerier even when they were presented inverted rather than upright. The increased eeriness for inverted faces was a notable finding, particularly as it was not mirrored in the results for how strongly participants felt fear when they were looking at the images either inverted or upright. While this may have been a result of the unusual qualities of the face blends, it may also have been because participants were unused to looking at faces when presented upside-down so the unusual nature of the task primed them to report the higher ratings of eeriness. An area for future research could be to present some of the typically 'eerie' NHFs such as those identified in Phase 1 in different orientations, and explore the extent to which this affects participant's tendency to report eeriness compared to ratings of human-likeness and strangeness.

There was no interaction between match-mismatch and orientation as mismatched faces were always eerier than matched ones and inverted faces were always eerier than upright ones. The results also supported MacDorman's (2009) theory that more extreme mismatches lead to more negative assessments, as the blend types where positive faces were blended with negative eyes were consistently rated as the most eerie, but these eeriness ratings were still not found to be particularly strong in terms of the scale used to rate eeriness. Four different negative emotions were included so it was also possible to identify which combinations resulted in the eeriest blends of individual emotions. It was

found that the *happy face, fearful eyes* and *happy face, angry eyes* were the eeriest blends in both orientations. A notable finding was that these results did not support the idea, grounded in Tinwell et al's 2013 research, that the eeriest faces would be those where the strong emotions were shown in the surrounding face, with a neutral expression in the eyes as a parallel to their 'expression lack' condition. None of the neutral ('dead') eye blends were rated as particularly eerie, and all were less eerie than the *happy face, angry eyes blend*. Considering all of the blends where neutral eyes were present, the blend with neutral eyes in *happy* and *angry* faces were eeriest (2.9) but that was still less eerie than the *happy face, angry eyes* blend. This finding can be interpreted with reference to Ekman et al's (2005) theory of leaked expressions, where witnessing someone who was attempting to conceal a negative expression by covering it with a positive one would induce a sense of distrust and unease in the observer. These extremely mismatched blends were presenting just that scenario, with something that looked like an attempt to conceal a fearful expression being found to be the most eerie of all. This also links into the theories of affinity and trust that were hypothesised to be influential in the UVE.

Gray and Wegner (2012) found that when an agent's appearance sets up an expectation that they should be able to experience human emotions yet their actual ability does not satisfy that expectation, that acceptability is reduced and unease is increased. In this way, a disconnect between two emotional expressions could set up a similar conflict and cause feelings of unease.

Finally, it is useful to consider the 2012 translation of Mori's (1970) original article, as this time, 'affinity' was used in the place of the original term of 'familiarity'. This subtle difference does lead to a new interpretation of the theory as affinity speaks more to the role of interaction and the ability to identify with a NHA. It can be seen how being unable

to form a connection with an NHA because of the conflicting signals given in incongruent expressions could result in an unsettling loss of affinity.

In addition to the key finding that mismatched faces elicit eeriness, two additional notable findings from Study 5 related to the accuracy of emotion identification, and participant's ability to categorise the different emotional blends. The ability to accurately identify emotions in the matched blends were recognised more accurately when presented upright, but the mismatched blends were recognised more accurately when presented inverted. Looking at all of the *positive face, negative eye blends*, the correct identification of emotion was particularly based on fear or anger in the eyes for the eeriest blends rather than the happy, smiling face component, which was used most for the disgusted, neutral or sad eye blends. These findings counter the happiness superiority effect proposed by Calvo and Lundqvist (2008) so it may be that there are some specific aspects of those particular blends that serves to over-ride the expected pattern where happiness would be identified, and instead are perceived in terms of the negative part of the face.

In terms of categorical perception theory and considering the two eeriest blends, the one hardest to categorise was the happy face with the fearful eyes, but none of the blends were found to be particularly difficult to categorise so this does not support Ramey's (2005) proposal that a source of the UVE may be ambiguity at categorical boundaries. In interpreting the results of this phase, it is important to consider whether they could have resulted from broader perceptual mechanisms governing emotional responses to faces. One additional explanation for the findings may be that certain types of blend are going to be more likely to give negative or threatening signals. Tipples (2007) found that faces with wide eyes and open mouths were most likely to be indicative of threat than other combinations of features. The *happy face, fearful eyes* blend that was

found to be the eeriest of the mismatches does present a wide-eyed appearance and while not all of the smiles are not indicative of open mouths, they are certainly unusual in the context of a wide-eyed expression. The importance of the wide-open eyes in this study also relates back to and provides support for Seyama and Nagayama's (2009) findings that enlarged eyes contribute to the UVE.

However, the research was not without its challenges, and some reflections on the methodologies used will now be briefly considered. This section will reflect on the methods used in this thesis and assess aspects of the methodology that may have contributed to the findings. A mixed methods approach was taken across the three research phases, with the later experimental phases informed by an initial exploration of a small number of stimuli which had been identified as potential candidates for eliciting an UVE. The manual coding approach used in Phase 1 did allow the experimenter to absorb the comments in detail and to gain a subjective understanding of the material which informed the design of the themes that were used to categorise the comments, but a disadvantage of this approach is that by reducing the free-text data to frequencies of coded quantitative observations, some of the richness of the original source will inevitably be lost. However, this approach was an appropriate method for an area that was relatively unexplored when the research was initiated as it meant the research topics for Phases 2 and 3 were able to emerge from an undirected exploration and to avoid the research being driven by taken for granted assumptions about the nature of the UVE.

In terms of aspects of the methodology which may have impacted on the findings of this thesis, the main area of concern relates to the morph images which were used in Phase 2. This has been discussed in detail in the General Discussion of Phase 2 but this study would have been improved considerably by using more morph stages rather than just the three plus the starting and end point images. Other studies using morphed

images have taken stills every 5%, 10% or 12.5% rather than the 25% used in this study and having a finer-grained view of the non-human to human scale would have allowed for a greater insight into the nature of the UVE. Technological improvements have also been in morphing tools, and more powerful ones are now available which would avoid some of the artefacts and issues found in the moped faces. It is now possible to create much smoother transitions and to apply local masking to areas where features do not always morph cleanly into each other, meaning that it would be possible to generate better stimuli which preserve the basic premise of a non-human to human transition but do so in a way which does not appear quite so obvious as a manipulated face.

Finally, the UVE is one which is hard to inspect in isolation from a broader societal and technological context as the ability to depict virtual humans continues to improve rapidly. In recent years there have been considerable developments in this area, such as improved software tools for manipulating images, and also in the creative environment where the virtual humans that are being designed for games and film continue to become more realistic and believable. If people adapt to NHAs through real-world exposure, research into the UVE risks falling behind the exposure participants may have experienced. It will be an on-going challenge to create stimuli that have the potential to reliably elicit a UVE. Given the long-term time frame of a research project of this nature, stimuli that may have been strong UV candidates at the development stage may no longer be at the eerie edge of realism by the time they come to be tested.

The findings across each of the research phases have been considered in terms of the underlying theories and research in each area. This discussion will now move back to the question of the UVE and discuss the extent to which this thesis has contributed to the development of techniques for research in this area.

6.2: Implications for the Uncanny Valley Theory

The UV is no longer considered an obscure concept and it is widely used in popular culture. Mainstream media now also make reference to it while discussing topics relating to robots or new technologies (e.g. Devlin (2015)). While this thesis has been in development, the academic landscape around the UV has changed dramatically and the effect is now seen as a reasonable area for serious enquiry. The original aim of this thesis was to pinpoint a single explanation for why the UVE occurs. The initial exploration phases indicated quickly that this was beyond the scope of a single research project so detailed attention was turned to understanding more about the specific domain of NHFs, how they are perceived and whether it would be possible to identify individual attributes responsible for the effect.

Looking back at Mori's (1970) original graph and revisiting his idea that increasing human-likeness would eventually cause the rejection of entities which came too close to human, what has this thesis been able to contribute to an understanding of that theory? Firstly, it has empirically demonstrated that there is a relationship between human-likeness, strangeness and eeriness and that the location of the eeriness peaks were found to be closer to the midpoint between non-human and human, but it has confirmed that there are entities that will be negatively evaluated as they increase in human-likeness. Secondly, it has looked in detail at how we perceive faces as they become more human-like, and contributed to an understanding of how NHFs are processed in the findings of Study 4 which found that NHFs are indeed subject to the effects of inversion, even more so than human faces. This thesis also developed Mori's idea by asking whether the effect was solely the result of a transition between robot and human or whether it would also appear when other types of non-humans were humanised. Study 4 examined animals, dolls and statues alongside robots and only the dolls showed a consistent *decrease* in

eeriness as human-likeness increased. The animals, robots and statues followed the expected pattern of a peak of eeriness part way along their transition to human. Dolls are sometimes described as eerie in their un-morphed state (Eberle (2009)) in a way that animals, robots and statues are not, so the addition of humanising features through morphing may have been caused them to follow a different trajectory through the valley. The dolls were also the only category where some of the matched humans were babies rather than adult faces so it may have been the transformation from doll to infant, rather than doll to human, that caused them to appear cuter, more appealing and therefore less eerie. Finally, differences in eeriness were found where mismatched emotional expressions did elicit significantly higher ratings of eeriness in participants, when compared to the original unmodified expressions. It was found that the strongest of these responses occurred when a happy face was presented with fearful eyes. However, the differences between the face images were found to be small and so further research would be needed to confirm whether this finding would apply more generally to NHFs.

In terms of an approach to investigating the UV, another contribution has been in the development of a set of design principles which it is hoped will be of use for future research in this area. These were detailed in Table 1 in Section 3.1. To recap, they were designed to avoid the inherent circularity present in several of the early studies in this area where it was taken for granted that certain types of NHA would evoke an uncanny response, and that certain agents were eerie simply by virtue of being close to human. The five principles were designed to guide the systematic and rigorous collection of the important measurements relating to the UVE, allowing unambiguous and data-driven conclusions to be drawn. They describe a framework for studies where it is possible to systematically increase the human-likeness of an agent from a given non-human starting point through to a fully realised version, indistinguishable from a human being. The

dimension described in this thesis has been the presentation of a realistic and convincing human face, but these principles could equally be applied to realistic motion or voice. It was felt that only by exerting careful control over the transition between non-human and human, while measuring the effects of that gradual transition, would it be possible to be certain of the relationship between human-likeness and emotional response, and in doing so, to identify characteristics of the NHAs that were most closely associated with eeriness and then explore those in more detail. The guiding principles were designed to aid practical research into the UVE but it was actually found that these are still difficult to apply in practice as even with objective, baseline measurements of the key dependent variables it is hard to draw conclusions about the nature of the UVE when participant responses of eeriness were generally rather low. It is also still difficult to avoid attempting to interpret differences in human-likeness, strangeness and eeriness as a result of some stimuli seeming subjectively eerier than others. For example, this was found in Study 4, where participant ratings of eeriness for morphed faces varied according to the category of the initial images. The finding that dolls were initially rated as eerie and then their eeriness decreased as human-likeness increased could have been explained as a result of the doll images being inherently eerier than animals, dolls and statues, and therefore already residing within an uncanny valley, but this explanation would risk falling foul of just that type of circularity.

This thesis has presented an exploration of the UV: findings have been noted in terms of the valley theory itself, with reference to wider psychological literature on face perception and a practical framework has been suggested which could be useful for future research designs looking to explore new aspects of the theory. This discussion will now move on to reflecting on some of the methodological choices that were made in designing

these studies and then looking at potential areas for future exploration before closing with some concluding comments.

6.3: Future Directions

The aim of this thesis was to explore the UVE and as the discussions above have demonstrated, has been able to contribute new findings in this area. However, there are certainly other areas for research within this discipline that are still to be explored. This section will describe three areas of potential new research.

Firstly, there has been relatively little research carried out systematically exploring the UVE in embodied agents and while virtual agents and images are now relatively well understood, studies of artificial agents in the real world have tended to be small scale in nature or carried out as case studies with specific reference to the testing of a particular aspect of an individual android. One intriguing possibility for further research would be looking at the highly realistic ‘reborn’ dolls that are now available that mimic the appearance of human babies. Eberle (2009) carried out a comparison of a reborn doll compared to a more traditional and stylised doll and found that a quarter of their respondents described both of them unsettling, but the strongest rejections were found for the reborn doll. This would contribute to an understanding of the UVE as these dolls are created to be appealing copies of human babies, not to disquiet or unsettle, but they seemed to have the opposite effect for some participants. In terms of the research carried out in this thesis it would be interesting to explore whether this effect can be generalised beyond the single dolls used in Eberle’s study, and if so to explore whether it may be that this occurs as another variant of the mismatch hypotheses, where in this case the mismatch would be between the highly realistic appearance paired with the inherent stillness and lack of animation present in a doll.

The second suggestion for further research would look for evidence of cultural differences in the perception of mismatched expressions to explore whether they are universally eerie, or whether there is a cultural influence on how easily a sense of uncanny can be elicited when happy faces and paired with fearful or angry eyes. Evidence for universal eeriness would support the idea that there may be an evolutionary component to the UVE while if culturally specific components can be found for the effect, it may be that exposure to particular types of face image may be responsible for triggering or flattening the effect.

Finally, this research is concluding at an exciting time in the world of virtual humans as rapid advances in computer technology mean that encounters with virtual agents are now commonplace and it is becoming increasingly difficult to distinguish digital actors from virtual ones. Journalists are now more likely to ask whether the latest film or game crosses the UV than to question whether the valley exists, as exemplified by Perry (2014) who reviewed recent innovations in this area in interviews with digital artists, film directors and two experts on the UVE (Tinwell and Saygin). The article considered the rise of 'digital actors', and looked at the challenges that will be presented if they do become commonplace over the next few years. Photorealistic actors such as Digital Ira (Debevec, 2012) are rendered at high levels of detail and have the ability to present a growing repertoire of convincing emotional expressions, and may be used in future to act alongside or even replace conventional actors in games and films. At the moment the actors are limited to short set pieces and the real challenge will be in making them portray convincing and emotionally engaging narratives in film, and even more so when they are the subjects for interaction in games. This challenge may present opportunities for developing the work in this thesis looking at mismatched emotional expressions in static images to exploring this further in animated faces, virtual agents or even beyond the

screen and into embodied agents. In addition, these initial findings about the perceptual mechanisms for NHFs could well be extended to virtual agents as they become more sophisticated, and could make important contributions to the ability to present convincing and relatable characters in these advanced computer graphics arenas. Digital Ira, like many convincing virtual actors, has been produced through motion capture of a real actor, whose appearance and emotion were then digitally recreated in such a way as to allow reproduction of realistic face movements and expressions. This technique has also been used to create virtual versions of well-known figures, often with unsettling or amusing results when the digital recreations are less than completely realistic. This suggests a new study as an area for future research concerning whether our familiarity with the actor would make us more sensitive to the flaws in their digital double. The digital recreation of an actor who is already well known presents a double challenge for the designer in that they not only have to design for a realistically human appearance but they need to reproduce the minute gestures and mannerisms that the audience will expect from the original. To date, no research has been published on the relationship between digital doubles and eeriness but it certainly seems to be a promising area for future enquiry as it would allow an exploration of the factors that influence the acceptability of different types of NHA. This thesis has suggested one such factor in expression mismatch, but by exploring a wider range of potential explanations it may be possible to achieve a greater depth of understanding of how NHAs are perceived.

6.4: Reflections on the Uncanny Valley Effect

This thesis has reviewed the prior theories and evidence for the UVE, and used these to inform the design of several studies which have considered the existence and nature of the valley from broad, qualitative explorations through to detailed and nuanced

experimental studies where variables have been carefully manipulated to allow conclusions to be drawn on some factors which appear to provide a compelling argument that certain types of faces are able to reliably elicit a sense of eeriness. The contributions that these explorations can make to general psychological knowledge, and specifically to the domain of face perception research, have been detailed above, but in closing it seems apposite to reflect on the idea of the uncanny valley itself and consider the extent to which it has real world applicability. The issue of methodological complexity in investigating the effect does make it a tricky one to pin down and given the eerie nature that characterises the effect, it can be easy to feel that research in this area is akin to chasing phantoms as they vanish into mist. Is there really an uncanny valley, or is it simply a convenient label to give to certain types of flawed representations of human-likeness to justify why they are rejected? My perspective, informed by research in this area and several years of near-constant exposure to a rapidly changing population of near-humans is that the effect is genuine, emotionally-driven and not yet completely understood. Androids, robots and CGI characters have developed to the point where they can delight, entertain and intrigue us, but they can still unsettle in that uniquely subtle way. I look forward to using the knowledge I have developed from this research study, and building on the findings in this thesis, to explore that fascinating effect in more detail.

References

- Baron-Cohen, S., Wheelwright, S., & Jolliffe, T. (1997). Is There a “Language of the Eyes?” Evidence from Normal Adults, and Adults with Autism or Asperger Syndrome. *Visual Cognition*, 4(3), 311–331.
- Bartneck, C., Kanda, T., Ishiguro, H., & Hagita, N. (2007). Is the uncanny valley an uncanny cliff? (pp. 368–373). Presented at the 16th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN 2007, Jeju, Korea: IEEE.
- Basit, T. (2003). Manual or electronic? The role of coding in qualitative data analysis. *Educational Research*, 45(2), 143–154.
- Bassili, J. N. (1979). Emotion recognition: the role of facial movement and the relative importance of upper and lower areas of the face. *Journal of personality and social psychology*, 37(11), 2049–2058. American Psychological Association.
- Bauer, R. M. (1986). The Cognitive Psychophysiology of Prosopagnosia. In H. D. Ellis, M. A. Jeeves, F. Newcombe, & A. Young (Eds.), *Aspects of Face Processing* (pp. 253–267). Dordrecht: Springer Netherlands.
- Becker-Asano, C., Ogawa, K., Nishio, S., & Ishiguro, H. (2010). Exploring the uncanny valley with Geminoid HI-1 in a real-world application (pp. 121–128). Presented at the IADIS Intl. Conf. Interfaces and Human Computer Interaction.
- Berthouze, L., & Berthouze, N. (2005). Affective Appraisal of Avatar Postures (pp. 19–20). Presented at the Proceedings of Views of the Uncanny Valley Workshop: IEEE-RAS International Conference on Humanoid Robots, Tsukuba, Japan.
- Brace, N. A., Hole, G. J., Kemp, R. I., Pike, G. E., Duuren, M. V., & Norgate, L. (2001). Developmental changes in the effect of inversion: Using a picture book to investigate face recognition. *Perception*, 30(1), 85–94.
- Breazeal, C. (2003). Emotion and sociable humanoid robots. *International Journal of Human-Computer Studies*, 59(1-2), 119–155.
- Breen, N., Caine, D., & Coltheart, M. (2000). Models of face recognition and delusional misidentification: A critical review. *Cognitive Neuropsychology*, 17(1-3), 55–71. Taylor & Francis.
- Brenton, H., Gillies, M., Ballin, D., & Chatting, D. (2005). The Uncanny Valley: Does it exist? Presented at the 20th National Conference on Artificial intelligence, Edinburgh.
- Bruce, V., & Young, A. W. (1986). Understanding face recognition. *British Journal of Psychology*, 77, 305–327.
- Bruce, V., & Young, A. W. (2000). *In the Eye of the Beholder: The Science of Face Perception*. Oxford University Press.
- Burleigh, T. J., Schoenherr, J. R., & Lacroix, G. L. (2013). Does the uncanny valley exist? An empirical test of the relationship between eeriness and the human likeness of digitally created faces. *Computers in Human Behaviour*, 29(3), 759–771.
- Calder, A. J. (1996). Categorical perception of morphed facial expressions. *Visual Cognition*, 3(2), 81–118. Taylor & Francis.
- Calder, A. J., Young, A. W., Keane, J., & Dean, M. (2000). Configural information in facial expression perception. *Journal of Experimental Psychology: Human Perception and Performance*, 26(2), 527. American Psychological Association.
- Calder, A., Young, A. W., Rowland, D., & Perrett, D. (1997). Computer-enhanced emotion in facial expressions, 264(1383), 919–925.
- Calvo, M. G., & Lundqvist, D. (2008). Facial expressions of emotion (KDEF): Identification under different display-duration conditions. *Behavior Research Methods*, 40(1), 109–

115.

- Cañamero, L. (2006). Did Garbo care about the uncanny valley? *Interaction Studies*, 7(3), 355–359.
- Carlston, D. E. (2013). On the Nature of Social Cognition: My Defining Moment. In *The Oxford Handbook of Social Cognition* (pp. 3–15). Oxford University Press.
- Chaminade, T., Hodgkins, J., & Kawato, M. (2005). Exploring the uncanny valley: behavioural and neuroimaging experiments (Vol. Tsukuba, Japan., pp. 21–23). Presented at the Proceedings of Views of the Uncanny Valley Workshop: IEEE-RAS International Conference on Humanoid Robots.
- Chaminade, T., Hodgkins, J., & Kawato, M. (2007). Anthropomorphism influences perception of computer-animated characters' actions. *Social Cognitive and Affective Neuroscience*, 2(3), 206–216.
- Cheetham, M. M., Suter, P. P., & Jäncke, L. L. (2011). The human likeness dimension of the “uncanny valley hypothesis”: behavioral and functional MRI findings. *Frontiers in Human Neuroscience*, 5, 126–126.
- Cheetham, M., Pavlovic, I., Jordan, N., Suter, P., & Jancke, L. (2013). Category Processing and the human likeness dimension of the Uncanny Valley Hypothesis: Eye-Tracking Data. *Frontiers in Psychology*, 4, 1–12.
- Clark-Carter, D. (1997). *Doing Quantitative Psychological Research*. Psychology Press.
- Cohen, L., & Cason, C. (2001). Do 7-month-old infants process independent features or facial configurations? *Infant and Child Development*, 10, 83–82.
- Cornes, K., Donnelly, N., Godwin, H., & Wenger, M. J. (2011). Perceptual and decisional factors influencing the discrimination of inversion in the Thatcher illusion. *Journal of Experimental Psychology: Human Perception and Performance*, 37(3), 645–668.
- Creed, C., & Beale, R. (2008). Psychological responses to simulated displays of mismatched emotional expressions. *Interacting with Computers*, 20, 225–239.
- Debevec, P. (2012). *Digital Ira*. Retrieved from <http://ict.usc.edu/prototypes/digital-ira/>
- Devlin, H. (2015, February 5). Rise of the robots: how long do we have until they take our jobs? Retrieved February 5, 2015, from <http://www.theguardian.com/technology/2015/feb/04/rise-robots-artificial-intelligence-computing-jobs>
- Diamond, R., & Carey, S. (1977). Developmental changes in the representation of faces. *Journal of Experimental Child Psychology*, 23(1), 1–22. Elsevier.
- Durand, K., Gallay, M., Seigneure, A., Robichon, F., & Baudouin, J.-Y. (2007). The development of facial emotion recognition: The role of configural information. *Journal of Experimental Child Psychology*, 97(1), 14–27.
- Eberle, S. G. (2009). Exploring the Uncanny Valley to Find the Edge of Play. *American Journal of Play*, 2(2), 1–28.
- Edwards, R., Honeycutt, J. M., & Zagacki, K. S. (1988). Imagined interaction as an element of social cognition. *Western Journal of Communication (includes Communication Reports)*, 52(1), 23–45. Taylor & Francis.
- Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, 6(3-4), 169–200. Taylor & Francis.
- Ekman, P., & Friesen, W. V. (1978). *Facial action coding system: A technique for the measurement of facial movement*. Palo Alto, CA: Consulting Psychologists Press.
- Ekman, P., Friesen, W. V., & O'Sullivan, M. (2005). Smiles When Lying. In P. Ekman & E. L. Rosenberg (Eds.), *What The Face Reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)* (2nd ed., pp. 201–216). Oxford University Press.

- Ellis, H. D., & Young, A. W. (1990). Accounting for delusional misidentifications. *The British Journal of Psychiatry*, 157(2), 239–248. The Royal College of Psychiatrists.
- Ellis, H. D., Young, A. W., Quayle, A. H., & De Pauw, K. W. (1997). Reduced autonomic responses to faces in Capgras delusion. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 264(1384), 1085–1092. The Royal Society.
- Fallshore, M., & Bartholow, J. (2003). Recognition of emotion from inverted schematic drawings of faces. *Perceptual and motor skills*, 96(1), 236–244.
- Farah, M. J., & Tanaka, J. W. (1995). What causes the face inversion effect? *Journal of Experimental Psychology: Human Perception and Performance*, 21(3), 628–634.
- Field, A. (2009). *Discovering Statistics using SPSS* (3rd ed.). SAGE.
- Freire, A., Lee, K., & Symons, L. A. (2000). The face-inversion effect as a deficit in the encoding of configural information: Direct evidence. *Perception*, 29(2), 159–170.
- Freud, S. (1919). The Uncanny. In *The Standard Edition of the Complete Psychological Works of Sigmund Freud* (Vol. 1917, pp. 217–256).
- Gee, F., Browne, W., & Kawamura, K. (2005). Uncanny valley Revisited (pp. 151–157). Presented at the IEEE International Workshop on Robots and Human Interactive Communication.
- Geller, T. (2008). Overcoming the uncanny valley. *IEEE Computer Graphics and Applications*, 11–17.
- Gotcher, J. M., & Edwards, R. (1990). Coping strategies of cancer patients: Actual communication and imagined interactions. *Health Communication*, 2(4), 255–266. Taylor & Francis.
- Gray, K., & Wegner, D. M. (2012). Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition*, 125(1), 125–130.
- Gray, S. (2006). *Exploring The Uncanny Valley*. Unpublished MSc Dissertation, The Open University.
- Green, R. D., MacDorman, K. F., Ho, C.-C., & Vasudevan, S. (2008). Sensitivity to the proportions of faces that vary in human likeness. *Computers in Human Behaviour*, 24(5), 2456–2474.
- Hanawalt, N. G. (1944). The role of the upper and the lower parts of the face as a basis for judging facial expressions: II. In posed expressions and “candid-camera” pictures. *The Journal of General Psychology*, 31(1), 23–36. Taylor & Francis.
- Hanson, D. (2003). *Investigating the Neural Basis of the Uncanny Valley*. Unpublished graduate research paper, The University of Texas at Dallas.
- Hanson, D., Olney, A., Zielke, M., & Pereira, I. (2005). Upending the Uncanny Valley (pp. 24–31). Presented at the 20th National Conference on Artificial intelligence, Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999.
- Hewson, C. (2003). Conducting research on the internet. *The Psychologist*, 16(6), 290–293.
- Hills, P. J., Ross, D. A., & Lewis, M. B. (2011). Attention misplaced: The role of diagnostic features in the face-inversion effect. *Journal of Experimental Psychology: Human Perception and Performance*, 37(5), 1396–1406.
- Ho, C.-C., MacDorman, K. F., & Pramono, Z. (2008). Human Emotion and the Uncanny Valley: A GLM, MDS, and Isomap Analysis of Robot Video Ratings. Presented at the 3rd ACM/IEEE international Conference on Human-Robot Interaction.
- Hole, G., & Bourne, V. (2010). *Face processing: Psychological, Neuropsychological, and Applied Perspectives*. Oxford University Press.
- Honeycutt, J. M., Zagacki, K. S., & Edwards, R. (1990). Imagined interaction and interpersonal communication. *Communication Reports*, 3(1), 1–8. Taylor & Francis.

- Horswill, M. S., & Coster, M. E. (2001). User-controlled photographic animations, photograph-based questions, and questionnaires: Three Internet-based instruments for measuring drivers' risk-taking behavior. *Behavior Research Methods, Instruments and Computers*, 33(1), 46–58.
- Hugengerg, K., & Wilson, J. P. (2013). Faces are central to social cognition. In D. E. Carlston (Ed.), *The Oxford Handbook of Social Cognition* (pp. 167–193). Oxford University Press.
- Ishiguro, H. (2005). Lateral Inhibition Hypothesis for Uncanny Valley (pp. 14–15). Presented at the Proceedings of Views of the Uncanny Valley Workshop: IEEE-RAS International Conference on Humanoid Robots, Tsukuba, Japan.
- Jentsch, E. (1906). On the Psychology of the Uncanny. *Angelaki*, 2, 7–16.
- Kang, M. (2009). The ambivalent power of the robot. *Antennae*, (9), 47–59.
- Kanwisher, N., & Moscovitch, M. (2000). The Cognitive Neuropsychology of Face Processing. *Cognitive Neuropsychology*, 17(1/2/3), 1–11. Psychology Pr.
- Kemp, S. (2004). *Future Face: Image, Identity, Innovation*. Profile Books.
- Keysers, C., & Gazzola, V. (2005). The neural basis of social cognitions and their responses to non-human agents (pp. 24–26). Presented at the Proceedings of Views of the Uncanny Valley Workshop: IEEE-RAS International Conference on Humanoid Robots, Tsukuba, Japan.
- Koda, T., & Ruttkay, Z. (2014). Eloquence of eyes and mouth of virtual agents: cultural study of facial expression perception. *AI & Society*.
- Kunda, Z. (1999). *Social Cognition*. MIT Press.
- Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallam, M., & Smoot, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin*, 126(3), 390–423. American Psychological Association.
- Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition and Emotion*, 24(8), 1377–1388.
- Leander, N. P., Chartrand, T. L., & Bargh, J. A. (2012). You Give Me the Chills: Embodied Reactions to Inappropriate Amounts of Behavioral Mimicry. *Psychological Science*, 23(7), 772–779.
- Leite, I., Pereira, A., Mascarenhas, S., Martinho, C., Prada, R., & Paiva, A. (2013). The influence of empathy in human–robot relations. *Journal of Human Computer Studies*, 71(3), 250–260. Elsevier.
- Lewkowicz, D. J., & Ghazanfar, A. A. (2011). The development of the uncanny valley in infants. *Developmental Psychobiology*, 54(2), 124–132.
- Lipp, O. V., Price, S. M., & Tellegen, C. L. (2009). No effect of inversion on attentional and affective processing of facial expressions. *Emotion*, 9(2), 248–259.
- Little, A. C., Jones, B. C., & DeBruine, L. M. (2011). The many faces of research on face perception. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1571), 1634–1637.
- Little, A. C., Penton-Voak, I. S., Burt, D. M., & Perrett, D. I. (2002). Evolution and individual differences in the perception of attractiveness: How cyclic hormonal changes and self-perceived attractiveness influence female preferences for male faces. In G. Rhodes & L. A. Zebrowitz (Eds.), *Facial attractiveness: Evolutionary, cognitive, and social perspectives*. (pp. 59–90). Westport, CT: Ablex Publishing.
- Looser, C., & Wheatley, T. (2010). The Tipping Point of Animacy: How, When, and Where We Perceive Life in a Face. *Psychological Science*, 21(12), 1854–1862.
- Lundqvist, D., Esteves, F., & Ohman, A. (1999). The face of wrath: Critical features for

- conveying facial threat. *Cognition and Emotion*, 13, 691-711.
- Lundqvist, D., Esteves, F., & Ohman, A. (2004). The face of wrath: The role of features and configurations in conveying social threat. *Cognition and Emotion*, 18, 161-182.
- MacDorman, K. F. (2005a). Androids as experimental apparatus: Why is there an uncanny valley and can we exploit it? (pp. 108-118). Presented at the CogSci Workshop Toward Social Mechanisms of Android Science, Stresa, Italy.
- MacDorman, K. F. (2005b). Memento Mori: Are Humanlike Robots Uncanny Because They Remind Us Of Death? (pp. 16-18). Presented at the Proceedings of Views of the Uncanny Valley Workshop: IEEE-RAS International Conference on Humanoid Robots, Tsukuba, Japan.
- MacDorman, K. F. (2005c). Mortality salience and the uncanny valley (pp. 399-405). Presented at the 5th IEEE RAS Conference on Humanoid Robotics.
- MacDorman, K. F. (2006). Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: An exploration of the uncanny valley. (pp. 26-29). Presented at the ICCSCogSci- Long Symposium Toward Social Mechanisms of Android Science, Vancouver, Canada.
- MacDorman, K. F., Green, R. D., Ho, C.-C., & Koch, C. T. (2009). Too real for comfort? Uncanny responses to computer generated faces. *Computers in Human Behaviour*, 25, 695-710.
- MacDorman, K. F., Srinivas, P., & Patel, H. (2013). The uncanny valley does not interfere with level 1 visual perspective taking. *Computers in Human Behaviour*, 29(4), 1671-1685. Elsevier.
- Martin, D., Slessor, G., Allen, R., Phillips, L. H., & Darling, S. (2012). Processing orientation and emotion recognition. *Emotion*, 12(1), 39-43.
- Mathur, M. B., & Reichling, D. B. (2009). An uncanny game of trust: Social trustworthiness of robots inferred from subtle anthropomorphic facial cues (pp. 313-314). Presented at the Proceedings of the 4th ACM/IEEE international conference on Human robot interaction.
- Matsuda, Y. T., Okamoto, Y., Ida, M., Okanoya, K., & Myowa-Yamakoshi, M. (2012). Infants prefer the faces of strangers or mothers to morphed faces: an uncanny valley between social novelty and familiarity. *Biology Letters*, 8(5), 725-728.
- Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *TRENDS in Cognitive Sciences*, 6(6), 255-260.
- Mauss, I. B., & Robinson, M. D. (2009). Measures of emotion: A review. *Cognition and Emotion*, 23(2), 209-237.
- McCann, R. M., & Honeycutt, J. M. (2006). A Cross-Cultural Analysis of Imagined Interactions. *Human communication research*, 32(3), 274-301.
- McDonnell, R. (2010). Face reality: investigating the Uncanny Valley for virtual faces. Presented at the ACM SIGGRAPH ASIA 2010 Sketches.
- Moosa, M. M., & Minhaz Ud-Dean, S. M. (2010). Danger avoidance: an evolutionary explanation of the uncanny valley. *Biological Theory*, 5(1), 12-14. MIT Press.
- Mori, M. (1970). The Uncanny Valley. *Energy*, 7(4), 33-35.
- Mori, M. (2005). On uncanny valley (p. 2). Presented at the Proceedings of Views of the Uncanny Valley Workshop: IEEE-RAS International Conference on Humanoid Robots, Tsukuba, Japan.
- Mori, M., MacDorman, K. F., & Kageki, N. (2012). The Uncanny Valley. *Robotics & Automation Magazine, IEEE*, 19(2), 98-100. IEEE.
- Navarrete, C. D., & Fessler, D. M. (2005). Normative Bias and Adaptive Challenges: A Relational Approach to Coalitional Psychology and a Critique of Terror Management

- Theory. *Evolutionary Psychology*, 3, 297–325.
- Pellicano, E., & Rhodes, G. (2003). Holistic processing of faces in preschool children and adults. *Psychological Science*, 14(6), 618–622. SAGE Publications.
- Perry, T. (2014). Leaving the uncanny valley behind. *Spectrum, IEEE*, 51(6), 48–53. IEEE.
- Peterson, M., & Rhodes, G. (2003). *Perception of Faces, Objects and Scenes*. Oxford University Press.
- Phillips, E., & Pugh, D. S. (1999). *How to Get a PhD: a Handbook for Students and Their Supervisors* (3rd ed.). Open University Press.
- Plantec, P. (2007). Digital Eye: Crossing the Great Uncanny Valley. *vfxworld.com*. Retrieved September 15, 2008, from <http://vfxworld.com/?atype=articles&id=3494>
- Pollick, F. (2005). The Search for the Uncanny Valley (pp. 4–7). Presented at the Proceedings of Views of the Uncanny Valley Workshop: IEEE-RAS International Conference on Humanoid Robots, Tsukuba, Japan.
- Pollick, F. (2009). In Search of the Uncanny Valley. In P. Daras & O. M. Ibarra (Eds.), *Lecture notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, UC Media* (pp. 69–78). Venice: Springer.
- Pollick, F., Paterson, H. M., Bruderlin, A., & Sanford, A. J. (2001). Perceiving affect from arm movement. *Cognition*, 82(2), B51–B61. Elsevier.
- Prkachin, G. C. (2003). The effects of orientation on detection and identification of facial expressions of emotion. *British Journal of Psychology*, 94(1), 45–62. Wiley Online Library.
- Quadflieg, S., & Macrae, C. N. (2010). Perceiving People. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of Social Psychology* (pp. 428–428). John Wiley & Sons.
- Quinn, K. A., & Macrae, C. N. (2011). The face and person perception: Insights from social cognition. *British Journal of Psychology*, 102(4), 849–867.
- Ramey, C. (2005). An Inventory of Reported Characteristics for Home Computers, Robots, and Human Beings: Applications for Android Science and the Uncanny Valley (pp. 8–13). Presented at the Proceedings of Views of the Uncanny Valley Workshop: IEEE-RAS International Conference on Humanoid Robots, Tsukuba, Japan.
- Reeves, M. (2011). *Code of Human Research Ethics* (pp. 1–44). The British Psychological Society.
- Reichardt, J. (1978). *Robots: Fact, Fiction and Prediction*. London: Thames and Hudson.
- Reips, U.-D. (2007). The methodology of Internet-based experiments. In A. Joinson, K. Y. A. McKenna, T. Postmes, & U.-D. Reips (Eds.), *Oxford Handbook of Internet Psychology* (pp. 1–19). Oxford University Press.
- Rhodes, G., Brake, S., & Atkinson, A. P. (1993). What's lost in inverted faces? *Cognition*, (47), 25–57.
- Riek, L., & Robinson, P. (2008). Real-Time Empathy: Facial Mimicry on a Robot (pp. 5–11). Presented at the 6th IEEE-RAS International Conference on Humanoid Robots.
- Rozin, P., & Fallon, A. (1987). A perspective on disgust. *Psychological review*, 94(1), 23–41.
- Russell, J. (1994). Is there universal recognition of emotion from facial expressions? A review of the cross-cultural studies. *Psychological Bulletin*, 115, 102–141.
- Sapsford, R. (2006). *Survey Research*. SAGE.
- Saygin, A. P., Chaminade, T., & Ishiguro, H. (2010). The Perception of Humans and Robots: Uncanny Hills in Parietal Cortex. In S. Ohlsson & R. Catrambone (Eds.), (pp. 2716–2720). Presented at the Proceedings of the 32nd Annual Conference of the Cognitive Science Society, Austin, Texas.
- Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J., & Frith, C. (2012). The thing that should not be: predictive coding and the uncanny valley in perceiving human and

- humanoid robot actions. *Social Cognitive and Affective Neuroscience*, 7(4), 413–422.
- Schneider, E., Wang, Y., & Yang, S. (2007). Exploring the Uncanny Valley with Japanese Video Game Characters (pp. 546–549). Presented at the DiGRA International Conference: Situated Play, Tokyo, Japan.
- Schwaninger, A., Lobmaier, J. S., Wallraven, C., & Collishaw, S. (2009). Two routes to face perception: evidence from psychophysics and computational modeling. *Cognitive Science*, 33(8), 1413–1440.
- Searcy, J., & Bartlett, J. (1996). Inversion and processing component and spatial-relational information in faces. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 904–915.
- Sekuler, A. B., Gaspar, C. M., Gold, J. M., & Bennett, P. J. (2004). Inversion Leads to Quantitative, Not Qualitative, Changes in Face Processing. *Current Biology*, 14(5), 391–396.
- Seyama, J., & Nagayama, R. S. (2007). The Uncanny Valley: Effect of Realism on the Impression of Artificial Human Faces. *Presence: Teleoperators & Virtual Environments*, 16(4), 337–351.
- Seyama, J., & Nagayama, R. S. (2009). Probing the uncanny valley with the eye size aftereffect. *Presence: Teleoperators & Virtual Environments*, 18(5), 321–339. MIT Press.
- Shimada, M., Minato, T., Itakura, S., & Ishiguro, H. (2007). Uncanny Valley of Androids and Its Lateral Inhibition Hypothesis (pp. 374–379). Presented at the 16th IEEE International Conference on Robot & Human Interactive Communication, Jeju, Korea.
- Slater, M. (2002). Presence and The Sixth Sense. *Presence: Teleoperators & Virtual Environments*, 11(4), 435–439.
- Slater, M., Brogni, A., & Steed, A. (2003). Physiological Responses to Breaks in Presence: A Pilot Study. Presented at the 6th Annual International Workshop on Presence, Aalborg, Denmark.
- Steckenfinger, S. A., & Ghazanfar, A. A. (2009). Monkey visual behavior falls into the uncanny valley. *Proceedings of the National Academy of Sciences*, 106(43), 18362–18366. National Acad Sciences.
- Swift, B. (2006). Preparing Numerical Data. In R. Sapsford & V. Jupp (Eds.), *Data Collection and Analysis* (pp. 153–183). SAGE.
- Takehara, T., & Suzuki, N. (1997). Morphed images of basic emotional expressions: Ratings on Russell's bipolar field. *Perceptual and motor skills*, 85(3), 1003–1010. Ammons Scientific.
- Thompson, P. (1980). Margaret Thatcher: a new illusion. *Perception*, 9, 482–484.
- Tinwell, A. (2009). Uncanny as Usability Obstacle. *Online Communities and Social Computing*, 622–631. Springer.
- Tinwell, A., & Grimshaw, M. (2009). Bridging the uncanny: an impossible traverse? (pp. 66–73). Presented at the 13th International MindTrek Conference: Everyday Life in the Ubiquitous Era, ACM.
- Tinwell, A., Grimshaw, M., & Williams, A. (2010). Uncanny behaviour in survival horror games. *Journal of Gaming & Virtual Worlds*, 2(1), 3–25. Retrieved from <http://www.ingentaconnect.com/content/intellect/jgvw/2010/00000002/00000001/art00001>
- Tinwell, A., Grimshaw, M., & Williams, A. (2011a). The Uncanny Wall. *International Journal of Arts and Technology*, 4(3), 326–341.
- Tinwell, A., Grimshaw, M., Nabi, D. A., & Williams, A. (2011b). Facial expression of emotion and perception of the Uncanny Valley in virtual characters. *Computers in*

- Human Behaviour*, 27(2), 741–749.
- Tinwell, A., Nabi, D. A., & Charlton, J. P. (2013). Perception of psychopathy and the Uncanny Valley in virtual characters. *Computers in Human Behaviour*, 29(4), 1617–1625.
- Tipples, J. (2007). Wide eyes and an open mouth enhance facial threat. *Cognition and Emotion*, 21(3), 535–557.
- Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. *British Journal of Psychology*, 79(4), 471–491. Wiley Online Library.
- Wegner, D. M., & Vallacher, R. R. (1977). *Implicit psychology: An introduction to social cognition*. Oxford University Press.
- Woods, S. (2006). Exploring the design space of robots: Children's perspectives. *Interacting with Computers*, 18, 1390–1418.
- Yin, R. K. (1969). Looking at Upside-Down Faces. *Journal of Experimental Psychology*, 81(1), 141–145.
- Young, G. (2009). In what sense “familiar?” Examining experiential differences within pathologies of facial recognition. *Consciousness and Cognition*, 18(3), 628–638. Elsevier Inc.
- Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception*, 16(6), 747–759.
- Yovel, G. (2005). The Neural Basis of the Behavioural Face Inversion Effect. *Current Biology*, 15, 2256–2262.
- Yuki, M., Maddux, W. W., & Masuda, T. (2007). Are the windows to the soul the same in the East and West? Cultural differences in using the eyes and mouth as cues to recognize emotions in Japan and the United States. *Journal of Experimental Social Psychology*, 43(2), 303–311.
- Ziemke, T., & Lindblom, J. (2006). Some methodological issues in android science. *Interaction Studies*, 7(3), 339–342.
- Zlotowski, J., & Bartneck, C. (2013). The inversion effect in HRI: are robots perceived more like humans or objects? (pp. 365–372). Presented at the IEEE international conference on Human-robot interaction, Tokyo, Japan.

Appendices

Phase 1

Appendix 1a: Participant demographics for the imagined encounters study

| Age group | N | % |
|--------------|-----|------|
| 20 and under | 14 | 6.6 |
| 21-30 | 77 | 36.3 |
| 31-40 | 76 | 35.8 |
| 41-50 | 29 | 13.7 |
| 51-60 | 11 | 5.2 |
| 60+ | 5 | 2.4 |
| Total | 212 | 100 |

Table 20: Distribution of participant ages

| Gender | N | % |
|--------|-----|------|
| Female | 141 | 66.5 |
| Male | 71 | 33.5 |
| Total | 212 | 100 |

Table 21: Distribution of participant gender

| Country | N | % |
|----------------|-----|-------|
| Argentina | 1 | 0.5 |
| Australia | 3 | 1.4 |
| Belgium | 1 | 0.5 |
| Canada | 13 | 6.1 |
| Germany | 2 | 0.9 |
| Ireland | 1 | 0.5 |
| Japan | 1 | 0.5 |
| Netherlands | 3 | 1.4 |
| Norway | 2 | 0.9 |
| Poland | 1 | 0.5 |
| South Africa | 1 | 0.5 |
| United Kingdom | 109 | 51.4 |
| United States | 73 | 34.4 |
| No answer | 1 | 0.5 |
| Total | 212 | 100.0 |

Table 22: Participant location

Appendix 1b: Example survey webpages

Introduction

For the purposes of this study, I would like you to imagine that you are attending a robotics exhibition. You will be asked to carry out three different kinds of task relating to figures that you encounter while you are at the exhibition. I will then ask some questions about yourself, but these will only be used to analyse your answers by group characteristics, never to identify you personally.

Please give your answers honestly, and feel free to use as much detail as you wish.

There are no right or wrong answers to the tasks.

Click 'proceed' when you are ready to start.

The screenshot shows a web browser window titled 'Stephanie Lay's Research Web - Windows Internet Explorer provided by The Open University'. The address bar shows the URL: http://mcs-notes2.open.ac.uk/UW/nf/Study2_web?OpenForm&Seq=1. The page content is as follows:

Stephanie Lay's Research Web: Study 2

Introduction

For the purposes of this study, I would like you to imagine that you are attending a robotics exhibition. You will be asked to carry out three different kinds of task relating to figures that you encounter while you are at the exhibition. I will then ask some questions about yourself, but these will only be used to analyse your answers by group characteristics, never to identify you personally.

Please give your answers honestly, and feel free to use as much detail as you wish. There are no right or wrong answers to the tasks.

Click 'proceed' when you are ready to start.

Proceed

You have answered
0 of 26 questions:

Task 1: 0 of 5,
Task 2: 0 of 5,
Task 3: 0 of 15.

(0% complete.)

© Stephanie Lay 2010/11
This site uses Fertigo Pro by Jos Buivenga / exilibris. Lotus Domino code and hosting kindly donated by MCT-IT Development team

Task One

You arrive at the exhibition while the displays are still being set up. You start to look around while you are waiting for your colleagues to arrive. You see five figures that you would really like them to take a look at, but you know that they will be in a hurry when they get here.

How would you describe this figure so that your colleagues will be able to find it easily?

Your description (text box)

Stephanie Lay's Research Web - Windows Internet Explorer provided by The Open University

http://mcs-notes2.open.ac.uk/UV/UV.nsf/Study2_web?OpenForm&Seq=4


File Edit View Favorites Tools Help

Stephanie Lay's Research Web

Stephanie Lay's Research Web: Study 2

Task One

You arrive at the exhibition while the displays are still being set up. You start to look around while you are waiting for your colleagues to arrive. You see five figures that you would really like them to take a look at, but you know that they will be in a hurry when they get here. How would you describe this figure so that your colleagues will be able to find it easily?



Your description

Submit this response

You have answered 0 of 26 questions:
Task 1: 0 of 5,
Task 2: 0 of 5,
Task 3: 0 of 15.
(0% complete.)

© Stephanie Lay 2010/11
This site uses Fertigo Pro by Jos Buivenga / exjlibris. Lotus Domino code and hosting kindly donated by MCT-IT Development team

Task completed.

Thank you for those answers. They have not yet been saved, because there are more tasks to complete. Please note that the instructions for the next task will be different from the previous task, even if they are similar.

Click 'proceed' when you are ready to continue.

Stephanie Lay's Research Web - Windows Internet Explorer provided by The Open University

http://mcs-notes2.open.ac.uk/UV/UV.nsf/Study2_web?OpenForm&Seq=16

File Edit View Favorites Tools Help

Stephanie Lay's Research Web

Stephanie Lay's Research Web: Study 2

Task completed.

Thank you for those answers. They have not yet been saved, because there are more tasks to complete. Please note that the instructions for the next task will be different from the previous task, even if they are similar. Click 'proceed' when you are ready to continue.

Proceed

You have answered 5 of 26 questions:
Task 1: 5 of 5,
Task 2: 0 of 5,
Task 3: 0 of 15.
(19.2% complete.)

© Stephanie Lay 2010/11
This site uses Fertigo Pro by Jos Buivenga / exjlibris. Lotus Domino code and hosting kindly donated by MCT-IT Development team

Task Two

During the exhibition, you learn that these figures are all prototypes for a new household robot. A researcher for the company who is building the robot would like to know how people would feel about having each of these in their home.

Please type in the box below: how does the figure make you feel?


Your description (text box)

Stephanie Lay's Research Web: Study 2

Task Two

During the exhibition, you learn that these figures are all prototypes for a new household robot. A researcher for the company who is building the robot would like to know how people would feel about having each of these in their home.

Please type in the box below: how does the figure make you feel?



Your feelings

Submit this response

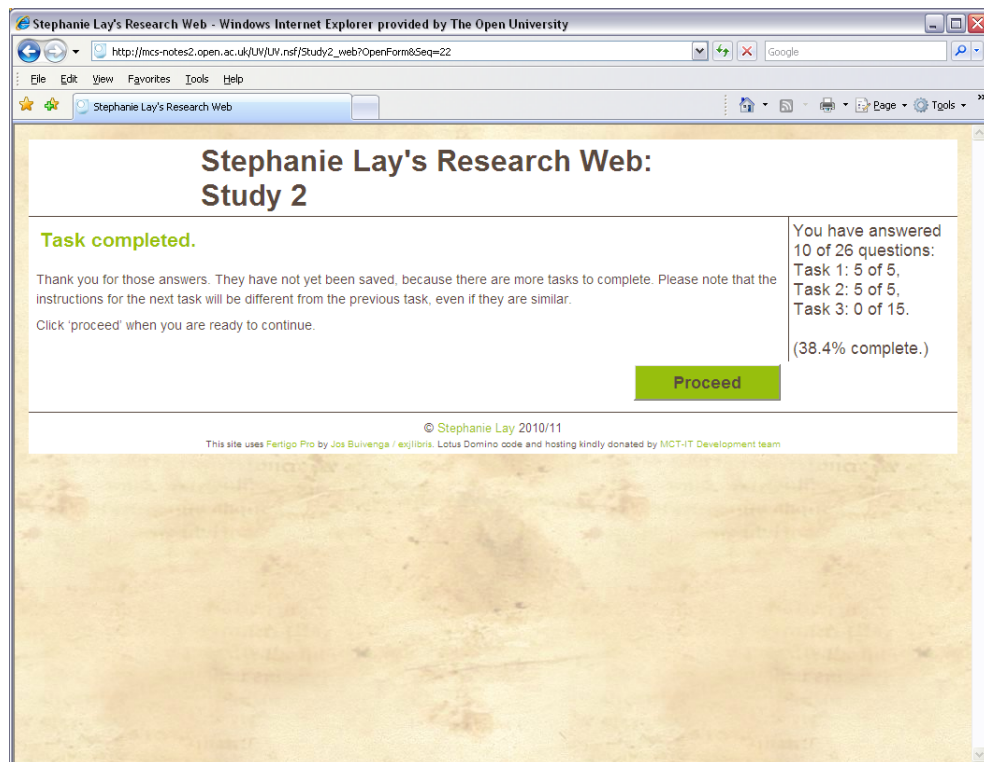
You have answered 5 of 26 questions:
 Task 1: 5 of 5,
 Task 2: 0 of 5,
 Task 3: 0 of 15.
 (19.2% complete.)

© Stephanie Lay 2010/11
 This site uses Fertigo Pro by Jos Buivenga / exilibris. Lotus Domino code and hosting kindly donated by MCT-IT Development team

Task completed.

Thank you for those answers. They have not yet been saved, because there are more tasks to complete. Please note that the instructions for the next task will be different from the previous task, even if they seem similar.

Click 'proceed' when you are ready to continue.



Task Three

At the end of the exhibition, you are asked to complete an evaluation of the five figures you have been interested in.

Please can you rate each of them on the following scales.

Note that the scales may vary for each question. The scale points are not numbered, as we want you to think about feelings rather than precise numbers.

Your rating:

Very strange ---> <--- Very familiar

Stephanie Lay's Research Web - Windows Internet Explorer provided by The Open University

http://mcs-notes2.open.ac.uk/JUV/JUV.nsf/Study2_web?OpenForm&Seq=23

File Edit View Favorites Tools Help

Stephanie Lay's Research Web


Stephanie Lay's Research Web: Study 2

Task Three

At the end of the exhibition, you are asked to complete an evaluation of the five figures you have been interested in.

Please can you rate each of them on the following scales.

Note that the scales may vary for each question. The scale points are not numbered, as we want you to think about feelings rather than precise numbers.



You have answered
10 of 26 questions:
Task 1: 5 of 5,
Task 2: 5 of 5,
Task 3: 0 of 15.
(38.4% complete.)

Your rating

<--- Very familiar Very strange --->

○ ○ ○ ○ ○ ○ ○ ○ ○ ○

Submit this response

© Stephanie Lay 2010/11
This site uses Fertigo Pro by Jos Buivenga / exjlibris. Lotus Domino code and hosting kindly donated by MCT-IT Development team

Stephanie Lay's Research Web - Windows Internet Explorer provided by The Open University

http://mcs-notes2.open.ac.uk/JUV/JUV.nsf/Study2_web?OpenForm&Seq=24

File Edit View Favorites Tools Help

Stephanie Lay's Research Web


Stephanie Lay's Research Web: Study 2

Task Three

At the end of the exhibition, you are asked to complete an evaluation of the five figures you have been interested in.

Please can you rate each of them on the following scales.

Note that the scales may vary for each question. The scale points are not numbered, as we want you to think about feelings rather than precise numbers.



You have answered
11 of 26 questions:
Task 1: 5 of 5,
Task 2: 5 of 5,
Task 3: 1 of 15.
(42.3% complete.)

Your rating

<--- Very mechanical Very human --->

○ ○ ○ ○ ● ○ ○ ○ ○ ○

Submit this response

© Stephanie Lay 2010/11
This site uses Fertigo Pro by Jos Buivenga / exjlibris. Lotus Domino code and hosting kindly donated by MCT-IT Development team

We are almost ready to save your answers.

- Your gender:

Male

Female

- Your age (years):
- Your location (City & Country):

One last thing... (optional)

If you found this study interesting, would you like to take part in more online research for Open University psychologists?

We are setting up a 'virtual participation panel' of people to participate in a wide variety of research projects. You can choose how often you take part, and you can leave the panel at any time. If you'd like to learn more, please leave your email address in the box below. Your email address will be used only for panel recruitment, and won't be used to identify the answers that you have just given.

- Enter your email address:

The screenshot shows a web browser window titled "Stephanie Lay's Research Web - Windows Internet Explorer provided by The Open University". The address bar shows the URL: http://mcs-notes2.open.ac.uk/UV/UV.nsf/Study2_web?OpenForm&Seq=39. The page content is as follows:

Stephanie Lay's Research Web: Study 2

Demographics

We are almost ready to save your answers.

In order to analyse whether the results vary according to participants' backgrounds, I'd like to know a little about you, please. This information will only be used to analyse answers by group characteristics, never to identify you personally. Please supply the following:

- Your gender:
 - ☐ Male
 - ☐ Female
- Your age (years):
- Your location (City & Country):
- Please select a country...

One last thing... (optional)

If you found this study interesting, would you like to take part in more online research for Open University psychologists?

We are setting up a 'virtual participation panel' of people to participate in a wide variety of research projects. You can choose how often you take part, and you can leave the panel at any time. If you'd like to learn more, please leave your email address in the box below. Your email address will be used only for panel recruitment, and won't be used to identify the answers that you have just given.

- Enter your email address:

Save all your answers

Appendix 1c: Frequency cross-tabulations of individual positive, negative and neutral emotion terms by face type

| Positive term | Artificial | Cluster 1 | Cluster 2 | Cluster 3 | Human | Grand Total |
|--------------------|------------|-----------|-----------|-----------|-------|-------------|
| comfortable | 10 | | 3 | 18 | 21 | 52 |
| Happiness happy | 15 | | | 18 | 7 | 40 |
| amused | 19 | | | | 2 | 21 |
| ok | 7 | 1 | 4 | 4 | 5 | 21 |
| at ease | 2 | | 1 | 9 | 6 | 18 |
| good | 3 | | 1 | 6 | 3 | 13 |
| comforted | 1 | | 1 | 5 | 3 | 10 |
| fine | 3 | 1 | | 5 | 1 | 10 |
| interested | 3 | 1 | 2 | 2 | | 8 |
| cheerful | 5 | | | | 1 | 6 |
| friendly | | | | 3 | 3 | 6 |
| safe | 2 | | | 2 | 2 | 6 |
| intrigued | 1 | | 2 | 1 | 1 | 5 |
| warm | | | | 2 | 3 | 5 |

| Positive term | Artificial | Cluster 1 | Cluster 2 | Cluster 3 | Human | Grand Total |
|----------------|------------|-----------|-----------|-----------|-------|-------------|
| confident | | | | 4 | | 4 |
| positive | 1 | | | 3 | | 4 |
| reassured | | | | 3 | 1 | 4 |
| acceptable | 1 | | 1 | 1 | | 3 |
| non-threatened | 1 | | | 1 | 1 | 3 |
| relaxed | | | | 1 | 2 | 3 |
| attracted | | | | 2 | | 2 |
| companionable | | | 1 | 1 | | 2 |
| funny | 2 | | | | | 2 |
| pleasant | 1 | | | 1 | | 2 |
| protective | 1 | | 1 | | | 2 |
| secure | 1 | | 1 | | | 2 |
| supported | | | | | 2 | 2 |
| charmed | 1 | | | | | 1 |
| compassionate | | | 1 | | | 1 |
| content | | | | 1 | | 1 |
| excited | | | | 1 | | 1 |
| fascinated | | | | 1 | | 1 |
| maternal | | 1 | | | | 1 |
| neighbourly | | | | 1 | | 1 |
| peaceful | | | | 1 | | 1 |
| pleased | | | | 1 | | 1 |
| trusting | | | | | 1 | 1 |
| Grand Total | 80 | 4 | 19 | 98 | 65 | 266 |

Table 23: Frequencies of key positive emotion terms for five face types, by descending frequency.

| Negative term | Artificial | Cluster 1 | Cluster 2 | Cluster 3 | Human | Grand Total |
|--------------------------|------------|-----------|-----------|-----------|-------|-------------|
| Creeped / creep / creepy | 8 | 33 | 10 | 4 | 8 | 63 |
| scared | 1 | 41 | 4 | 2 | 2 | 50 |
| uncomfortable | 3 | 11 | 23 | 4 | 9 | 50 |
| uneasy | 1 | 9 | 14 | 3 | 5 | 32 |
| disturbed/disturbing | 1 | 13 | 5 | 2 | 3 | 24 |
| unsettled | 4 | 4 | 8 | 4 | 1 | 21 |
| unnerved | 1 | 5 | 9 | | 2 | 17 |
| freaked/freaky | 2 | 10 | 1 | 1 | 2 | 16 |
| weird | 3 | 1 | 4 | 1 | | 9 |
| frightened/frightening | | 5 | 3 | | | 8 |

| Negative term | Artificial | Cluster 1 | Cluster 2 | Cluster 3 | Human | Grand Total |
|-----------------|------------|-----------|-----------|-----------|-------|-------------|
| scrutinised | 2 | | 6 | | | 8 |
| terrified | | 8 | | | | 8 |
| bad | | 4 | 1 | 1 | 1 | 7 |
| disconcerted | | 2 | 3 | 1 | 1 | 7 |
| sad | 2 | 2 | 2 | | 1 | 7 |
| nervous | | 2 | 3 | 1 | | 6 |
| strange | | | 1 | 1 | 4 | 6 |
| annoyed | 1 | | | | 4 | 5 |
| intimidated | 1 | | 2 | 1 | 1 | 5 |
| bored | 1 | 1 | | 1 | 1 | 4 |
| cold | 1 | | 3 | | | 4 |
| repulsed | 2 | 1 | 1 | | | 4 |
| spooked | | 2 | 2 | | | 4 |
| anxious | | 3 | | | | 3 |
| disgusted | | | | 2 | 1 | 3 |
| eerie | | 1 | 2 | | | 3 |
| threatened | | | 2 | 1 | | 3 |
| unpleasant | | 2 | 1 | | | 3 |
| worried | 1 | | 1 | | 1 | 3 |
| apathetic | 2 | | | | | 2 |
| apprehensive | | 1 | 1 | | | 2 |
| awful | | 2 | | | | 2 |
| bothered | | 2 | | | | 2 |
| cautious | | | | 1 | 1 | 2 |
| confused | | | 1 | | 1 | 2 |
| depressed | | | 1 | | 1 | 2 |
| disdainful | 1 | | | | 1 | 2 |
| distant | 1 | 1 | | | | 2 |
| embarrassed | | | 1 | 1 | | 2 |
| guilty | | | 1 | | 1 | 2 |
| jealous | | | 1 | 1 | | 2 |
| judged | | | 1 | | 1 | 2 |
| repelled | | | 1 | | 1 | 2 |
| suspicious | 1 | | 1 | | | 2 |
| wary | | | 1 | | 1 | 2 |
| alarmed/alarmed | | 1 | | | | 1 |
| angry | | 1 | | | | 1 |
| awkward | | | 1 | | | 1 |

| Negative term | Artificial | Cluster 1 | Cluster 2 | Cluster 3 | Human | Grand Total |
|---------------|------------|-----------|-----------|-----------|-------|-------------|
| concerned | | 1 | | | | 1 |
| conflicted | | | | | 1 | 1 |
| guarded | | | 1 | | | 1 |
| haunted | | 1 | | | | 1 |
| horrified | | 1 | | | | 1 |
| nightmarish | | 1 | | | | 1 |
| paranoid | | | 1 | | | 1 |
| perturbed | | 1 | | | | 1 |
| tense | | 1 | | | | 1 |
| unacceptable | | | | 1 | | 1 |
| Grand Total | 40 | 174 | 124 | 34 | 56 | 428 |

Table 24: Frequencies of key negative emotion terms for five face types, by descending frequency.

| Neutral term | Artificial | Cluster 1 | Cluster 2 | Cluster 3 | Human | Grand Total |
|------------------|------------|-----------|-----------|-----------|-------|-------------|
| neutral | 8 | 1 | 11 | 8 | 13 | 41 |
| calm | 1 | | | 7 | 2 | 10 |
| indifferent | 1 | | 4 | 1 | 3 | 9 |
| ambivalent | 2 | | 1 | 1 | 3 | 7 |
| curious | 4 | | 2 | | | 6 |
| normal | | | | 2 | 3 | 5 |
| mixed | 1 | | | | 1 | 2 |
| nostalgic | | | | | 2 | 2 |
| amazed | | | | 1 | | 1 |
| Calm/ curious | 1 | | | | | 1 |
| emotionless | | | 1 | | | 1 |
| maternal | 1 | | | | | 1 |
| puzzled | 1 | | | | | 1 |
| startled | | 1 | | | | 1 |
| Grand Total | 20 | 2 | 19 | 20 | 27 | 88 |

Table 25: Frequencies of key neutral emotion terms for five face types, by descending frequency.

| Term | Eeriness synonym |
|------------------------|------------------|
| creeped/creeps/creepy | Yes |
| disconcerted | Yes |
| disturbed/disturbing | Yes |
| eerie | Yes |
| freaked/freaky | Yes |
| frightened/frightening | Yes |

| Term | Eeriness synonym |
|-----------------|------------------|
| haunted | Yes |
| nervous | Yes |
| nightmarish | Yes |
| scared | Yes |
| spooked | Yes |
| uneasy | Yes |
| unnerved | Yes |
| unsettled | Yes |
| alarmed/alarmed | No |
| angry | No |
| annoyed | No |
| anxious | No |
| apathetic | No |
| apprehensive | No |
| awful | No |
| awkward | No |
| bad | No |
| bored | No |
| bothered | No |
| cautious | No |
| cold | No |
| concerned | No |
| conflicted | No |
| confused | No |
| depressed | No |
| disdainful | No |
| disgusted | No |
| distant | No |
| embarrassed | No |
| guarded | No |
| guilty | No |
| horrified | No |
| intimidated | No |
| jealous | No |
| judged | No |
| paranoid | No |
| perturbed | No |
| repelled | No |

| Term | Eeriness synonym |
|---------------|------------------|
| repulsed | No |
| sad | No |
| scrutinized | No |
| strange | No |
| suspicious | No |
| tense | No |
| terrified | No |
| threatened | No |
| unacceptable | No |
| uncomfortable | No |
| unpleasant | No |
| wary | No |
| weird | No |
| worried | No |

Table 26: Negative terms used to in descriptions of emotional reactions towards near-human agents, indicating whether they are synonymous with eeriness.

Phase 2

Appendix 2a: Participant demographics for rating study

| Sex | N | % |
|----------------|----------|----------|
| Female | 73 | 68.2 |
| Male | 32 | 29.9 |
| Refused | 2 | 1.9 |
| Total | 107 | 100 |
| Age | | |
| <21 | 2 | 1.9 |
| 21 - 25 | 7 | 6.5 |
| 26 - 30 | 22 | 20.6 |
| 31 - 35 | 27 | 25.2 |
| 36 - 40 | 13 | 12.1 |
| 41 - 45 | 7 | 6.5 |
| 46 - 50 | 7 | 6.5 |
| 51 - 55 | 5 | 4.7 |
| 56 - 60 | 10 | 9.3 |
| >60 | 5 | 4.7 |
| Refused | 2 | 1.9 |
| Total | 107 | 100 |
| Country | | |
| Germany | 1 | 0.9 |
| Ireland | 1 | 0.9 |
| Netherlands | 1 | 0.9 |
| Norway | 1 | 0.9 |
| United Kingdom | 86 | 80.4 |
| United States | 15 | 14.0 |
| Refused | 2 | 1.9 |
| Total | 107 | 100 |

Table 27: Participant demographics for rating study

Appendix 2b: Participant demographics for inversion study

| Gender | N | % |
|---------------|----------|----------|
| Female | 35 | 64.8 |
| Male | 19 | 35.2 |
| Total | 54 | 100 |

| Age | | |
|--------------|-----------|------------|
| 21 - 25 | 1 | 2.0 |
| 26 - 30 | 5 | 10.2 |
| 31 - 35 | 9 | 18.4 |
| 36 - 40 | 7 | 14.3 |
| 41 - 45 | 6 | 12.2 |
| 46 - 50 | 8 | 16.3 |
| 51 - 55 | 5 | 10.2 |
| 56 - 60 | 5 | 10.2 |
| >60 | 3 | 6.1 |
| Not known | 5 | 6.1 |
| Total | 54 | 100 |

Table 28: Participant demographics for inversion study

Appendix 2c: Image sources

This table displays the original source of the non-human and human images used to create the morph animations and morph stage stills. The experimental images are shown first followed by the foil images with the three training images shown at the end of the table.

| Experiment role | Image Category | Source Link |
|-----------------|----------------|---|
| Experimental | Animal | http://files.myopera.com/Bjork/albums/60620/doggie5.jpg |
| Experimental | Animal | http://files.myopera.com/Bjork/albums/60620/doggie5.jpg |
| Experimental | Animal | http://files.myopera.com/Bjork/albums/60620/doggie4.jpg |
| Experimental | Animal | http://files.myopera.com/Bjork/albums/60620/doggie4.jpg |
| Experimental | Animal | http://www.flickr.com/photos/aturkus/279404991/ |
| Experimental | Animal | http://www.flickr.com/photos/mein_arkengel/3994508518/ |
| Experimental | Doll | http://www.flickr.com/photos/a_mason/53722615/ |
| Experimental | Doll | http://www.flickr.com/photos/paramourphotos/4340952343/sizes/o/in/photostream/ |
| Experimental | Doll | http://www.flickr.com/photos/chantelbeam/5678044316/ |
| Experimental | Doll | http://www.flickr.com/photos/chantelbeam/5678044316/ |
| Experimental | Doll | http://www.flickr.com/photos/bergius/77312481/ |
| Experimental | Doll | http://www.flickr.com/photos/a4gpa/5703348740/sizes/o/in/photostream/ |
| Experimental | Robot | Taken from previous MSc study |
| Experimental | Robot | Taken from previous MSc study |
| Experimental | Robot | http://cache.boston.com/universal/sitegraphics/blogs/bigpicture/robots_03_04/r05_17159729.jpg |
| Experimental | Robot | http://www.flickr.com/photos/abolotnov/3373366268/sizes/ |



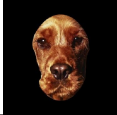





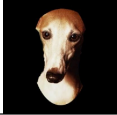

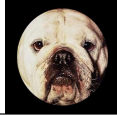
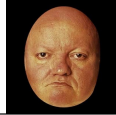
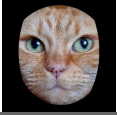
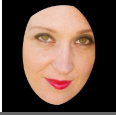
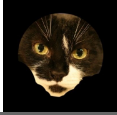
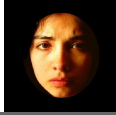
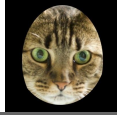


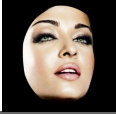



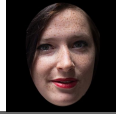










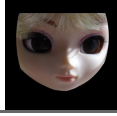












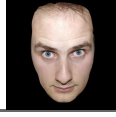









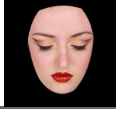
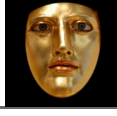



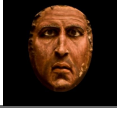








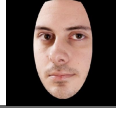
| Experiment role | Image Category | Source Link |
|-----------------|----------------|---|
| | | o/in/photostream/ |
| Experimental | Robot | http://www.kyosho.com/blog/manoi/log/file/マノイPF右見上げ.jpg |
| Experimental | Robot | http://www.flickr.com/photos/gregpc/3011194936/sizes/l/in/photostream/ |
| Experimental | Statue | http://www.flickr.com/photos/jeremybrooks/2055577531/ |
| Experimental | Statue | http://danceadvantage.net/wp-content/uploads/2010/11/SophiaEyesClosed-266x400.jpg |
| Experimental | Statue | http://www.flickr.com/photos/tentacles/2230946717/ |
| Experimental | Statue | http://www.flickr.com/photos/toddiestand/180314723/sizes/l/in/photostream/ |
| Experimental | Statue | http://www.flickr.com/photos/michalo/4473670045/ |
| Experimental | Statue | http://www.flickr.com/photos/buginbox/4940404337/ |
| Foil | Animal | http://creativeadvertisingworld.com/wp-content/uploads/2007/08/cesar-20dog.jpg |
| Foil | Animal | http://creativeadvertisingworld.com/wp-content/uploads/2007/08/cesar-20dog.jpg |
| Foil | Animal | http://creativeadvertisingworld.com/wp-content/uploads/2007/08/dog-20food.jpg |
| Foil | Animal | http://creativeadvertisingworld.com/wp-content/uploads/2007/08/dog-20food.jpg |
| Foil | Animal | http://creativeadvertisingworld.com/wp-content/uploads/2007/08/cesar-20dog-20food.jpg |
| Foil | Animal | http://creativeadvertisingworld.com/wp-content/uploads/2007/08/cesar-20dog-20food.jpg |
| Foil | Animal | http://creativeadvertisingworld.com/wp-content/uploads/2007/08/dog-20food-20cesar.jpg |
| Foil | Animal | http://creativeadvertisingworld.com/wp-content/uploads/2007/08/dog-20food-20cesar.jpg |
| Foil | Animal | http://www.flickr.com/photos/mrsenil/564080909/sizes/l/in/photostream/ |
| Foil | Animal | http://www.flickr.com/photos/59398710@N07/5987216885/in/set-72157627183268525 |
| Foil | Animal | http://www.flickr.com/photos/mydesignselfstudy/3144039153/sizes/o/in/photostream/ |
| Foil | Animal | http://www.flickr.com/photos/nile/red/162805718/sizes/o/in/photostream/ |
| Foil | Doll | http://www.flickr.com/photos/8137371@N05/3779284235/sizes/o/in/photostream/ |
| Foil | Doll | http://www.flickr.com/photos/8137371@N05/3779284235/sizes/o/in/photostream/ |
| Foil | Doll | http://www.brocobellevintage.co.uk/ekmps/shops/karenbrocshop/images/vtg-50s-60s-cutest-baby-doll-face-for-rag-doll-making-24-p.jpg |
| Foil | Doll | http://www.flickr.com/photos/edgarbarany/4387342149/siz |

| Experiment role | Image Category | Source Link |
|-----------------|----------------|---|
| | | es/o/in/photostream/ |
| Foil | Doll | http://fc05.deviantart.net/fs71/i/2010/263/1/8/pocahontas doll face repaint by orchideah-d2z5osu.jpg |
| Foil | Doll | http://www.flickr.com/photos/colinlogan/5234459648/sizes/o/in/photostream/ |
| Foil | Doll | http://www.flickr.com/photos/jeremybrooks/2246729079/sizes/o/in/photostream/ |
| Foil | Doll | http://www.flickr.com/photos/andresthor/3998076363/sizes/o/in/photostream/ |
| Foil | Doll | http://www.mannequinstore.com/busts/images/HB2/head.jpg |
| Foil | Doll | http://www.flickr.com/photos/mait/4300417797/sizes/o/in/photostream/ |
| Foil | Doll | http://www.flickr.com/photos/babelglyph/5617034509/ |
| Foil | Doll | http://www.flickr.com/photos/23680544@N07/3616010984 |
| Foil | Robot | http://www.fastfancydress.co.uk/templates/imagedirectory/robot%20mask%20lg.jpg |
| Foil | Robot | http://www.flickr.com/photos/mindonfire/2592282205/sizes/o/in/photostream/ |
| Foil | Robot | http://www.flickr.com/photos/jurvetson/2218864889/sizes/l/in/photostream/ |
| Foil | Robot | http://www.flickr.com/photos/pauldwaite/214450820/sizes/l/in/photostream/ |
| Foil | Robot | http://4.bp.blogspot.com/_EfNKxLWT6wA/S8CsQRdIfgl/AAAAAAAAAANs/2wCzoVy7WEM/s1600/irobot2.jpg |
| Foil | Robot | http://www.flickr.com/photos/kliefi/3071616767/sizes/o/in/photostream/ |
| Foil | Robot | http://teamblog.brainsunleashed.org/wordpress/wp-content/uploads/2011/02/Nexi.jpg |
| Foil | Robot | http://www.flickr.com/photos/foreverphoto/2188839612/sizes/l/in/photostream/ |
| Foil | Robot | http://www.indiatalkies.com/images/robot-head-7738m.jpg |
| Foil | Robot | http://www.flickr.com/photos/chiaralily/3639632097/sizes/l/in/photostream/ |
| Foil | Robot | http://www.wallpaperpimper.com/wallpaper/Computer/NEC/Gundam-World-Papero-Nec-1-8CLXUV6YS6-1280x1024.jpg |
| Foil | Robot | http://www.flickr.com/photos/alexis/1486824926/sizes/l/in/photostream/ |
| Foil | Statue | http://www.free-photos.biz/images/luxury/jewellery/face of statue of liberty.jpg |
| Foil | Statue | http://www.flickr.com/photos/alainbachellier/3131290814/sizes/l/in/photostream/ |
| Foil | Statue | http://4.bp.blogspot.com/_FzeX7fiyEGo/TAdESdxJlBl/AAAAAAAAAF0/1rC52UQhdeo/s1600/Kelly face hi res.jpg |

| Experiment role | Image Category | Source Link |
|-----------------|----------------|---|
| Foil | Statue | http://www.flickr.com/photos/pinkstockphotos/5263478657/sizes/l/in/photostream/ |
| Foil | Statue | http://www.flickr.com/photos/maynard/97495610/sizes/l/in/photostream/ |
| Foil | Statue | http://www.flickr.com/photos/psexypsychic/5065800931/sizes/l/in/photostream/ |
| Foil | Statue | http://www.flickr.com/photos/11287317@N04/2975805610/sizes/o/in/photostream/ |
| Foil | Statue | http://www.flickr.com/photos/jasonpier/5415215335/sizes/l/in/photostream/ |
| Foil | Statue | http://www.flickr.com/photos/fahrradfritze/2206736999/sizes/l/in/photostream/ |
| Foil | Statue | http://www.flickr.com/photos/hugo971/5734794202/sizes/l/in/photostream/ |
| Foil | Statue | http://ny-image2.etsy.com/il_fullxfull.64608138.jpg |
| Foil | Statue | http://www.flickr.com/photos/nihonbunka/2716143841/sizes/l/in/photostream/ |
| Training | Fruit | http://www.flickr.com/photos/patdavid/5568423570/sizes/l/in/photostream/ |
| Training | Fruit | http://www.flickr.com/photos/tomsaint/3171749726/sizes/l/in/photostream/ |
| Training | Fruit | http://www.flickr.com/photos/marcel030nl/2377272785/sizes/l/in/photostream/ |

Table 29: Original image source links for Phase 2

Appendix 2d: Foil and experimental images

| | Foil | | Experimental | | Foil | |
|--------|---|---|---|--|---|---|
| | Morph anchor: non human | Morph anchor: human | Morph anchor: non human | Morph anchor: human | Morph anchor: non human | Morph anchor: human |
| Animal |  |  |  |  |  |  |
| |  |  |  |  |  |  |
| |  |  |  |  |  |  |
| Doll |  |  |  |  |  |  |
| |  |  |  |  |  |  |
| |  |  |  |  |  |  |
| Robot |  |  |  |  |  |  |
| |  |  |  |  |  |  |
| |  |  |  |  |  |  |
| Statue |  |  |  |  |  |  |
| |  |  |  |  |  |  |
| |  |  |  |  |  |  |

Training Images



Table 30: All foil and experimental images used for Phase 2, by image category and non-human/human anchor.

Appendix 2e: Online study texts

Cognitive Psychology Rating Exercise

Thank you for your interest!

I'd like you to take part in a research activity investigating face perception. This makes up part of my PhD research and is supervised by academics from the Open University and Glasgow University.

HOW LONG WILL THE STUDY BE LIVE?

The study will run from July 22nd to August 21st 2011.

AIM OF THE RESEARCH ACTIVITY

This online activity involves looking at images of faces and giving your judgements about them on several different scales: this is to investigate psychological aspects of face perception.

WHAT DATA WILL BE COLLECTED?

I will collect the ratings that you give for each image. At the end of the activity I will ask for some basic demographic information so I can make sure that my conclusions are fair and not biased towards or against any particular group. This part is optional: you can still take part and choose not to give this information. I may also collect your email address at the end of the activity if you wish to join my mailing list to receive a short report of the results of the activity.

HOW WILL THE DATA BE COLLECTED?

The data will be collected from the responses you give in the online forms on the following pages.

CONFIDENTIALITY/ANONYMITY

The data I collect do not contain any personally sensitive information. If you choose to give any demographic information this will not be linked to the ratings you give in the

main part of the activity. This research is carried out in compliance with the Data Protection and Freedom of Information Acts.

TIME COMMITMENT

The activity has been tested and will require around 20 minutes to complete: this will take place in a single session.

WITHDRAWING FROM THE ACTIVITY

You may decide to stop taking part in the activity at any time without explanation. If you decide you no longer want to carry on, just close your web browser. Your responses up to that point will be excluded from the final analysis. If you complete the activity and decide at a later date that you would like to withdraw your responses, please contact me before the activity closing date and I will ensure they are securely deleted.

RISKS

There are no known risks for you in taking part in this activity.

COST, REIMBURSEMENT AND COMPENSATION

Your participation in this activity is voluntary and no reimbursement or compensation is offered for your time. However, you will be contributing to research in a novel area of psychology, and your participation is greatly appreciated.

FURTHER INFORMATION ABOUT THIS RESEARCH ACTIVITY

When you have completed the activity you will be able to read more about this research and how your participation in this activity has contribute to the overall project. There will also be an opportunity to leave your email address if you wish to receive a short report on the results of this activity.

CONTACTS

I (Stephanie Lay) will be glad to answer your questions about this study at any time. If you have a query that I am unable to answer, please contact one of my supervisors.

| Principle Investigator | Supervisor | Supervisor |
|--|--|--|
| Stephanie Lay | Nicky Brace | Graham Pike |
| Psychology Department | Psychology Department | Psychology Department |
| The Open University | The Open University | The Open University |
| Walton Hall | Walton Hall | Walton Hall |
| Milton Keynes | Milton Keynes | Milton Keynes |
| MK7 6AA | MK7 6AA | MK7 6AA |
| S.C.Lay@open.ac.uk | N.A.Brace@open.ac.uk | G.E.Pike@open.ac.uk |

Do you confirm that you are over 18 years of age, and that you accept these terms and conditions?

The screenshot shows a web browser window with the address bar displaying 'uncanny-valley.open.ac.uk/UV/UV.nsf/Study3_web_CLOSED?OpenForm&Seq=160'. The page title is 'Stephanie Lay's Research Web: Phase 2'. The content is organized into sections with green headers:

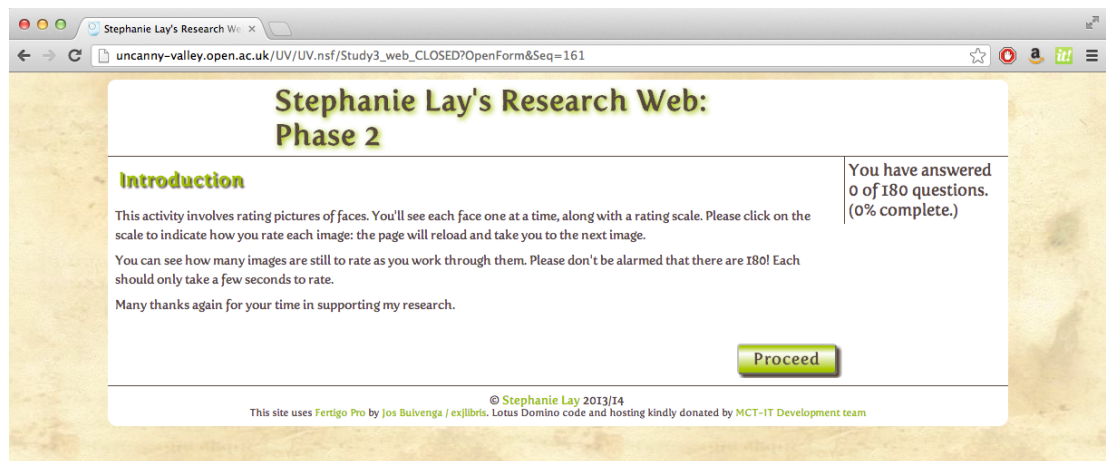
- Terms & Conditions**
- Cognitive Psychology Rating Exercise**
 - Thank you for your interest!
 - I'd like you to take part in a research activity investigating face perception. This makes up part of my PhD research and is supervised by academics from the Open University and Glasgow University.
- HOW LONG WILL THE STUDY BE LIVE?**
 - The study will run from July 22nd to August 21st 2011.
- AIM OF THE RESEARCH ACTIVITY**
 - This online activity involves looking at images of faces and giving your judgements about them on several different scales; this is to investigate psychological aspects of face perception.
- WHAT DATA WILL BE COLLECTED?**
 - I will collect the ratings that you give for each image. At the end of the activity I will ask for some basic demographic information so I can make sure that my conclusions are fair and not biased towards or against any particular group. This part is optional: you can still take part and choose not to give this information. I may also collect your email address at the end of the activity if you wish to join my mailing list to receive a short report of the results of the activity.
- HOW WILL THE DATA BE COLLECTED?**
 - The data will be collected from the responses you give in the online forms on the following pages.
- CONFIDENTIALITY/ANONYMITY**
 - The data I collect do not contain any personally sensitive information. If you choose to give any demographic information this will not be linked to the ratings you give in the main part of the activity. This research is carried out in compliance with the Data Protection and Freedom of Information Acts.
- TIME COMMITMENT**
 - The activity has been tested and will require around 20 minutes to complete: this will take place in a single session.
- WITHDRAWING FROM THE ACTIVITY**
 - You may decide to stop taking part in the activity at any time without explanation. If you decide you no longer want to carry on, just close your web browser. Your responses up to that point will be excluded from the final analysis. If you complete the activity and decide at a later date that you would like to withdraw your responses, please contact me before the activity closing date and I will ensure they are securely deleted.
- RISKS**
 - There are no known risks for you in taking part in this activity.
- COST, REIMBURSEMENT AND COMPENSATION**

Introduction

This activity involves rating pictures of faces. You'll see each face one at a time, along with a rating scale. Please click on the scale to indicate how you rate each image: the page will reload and take you to the next image.

You can see how many images are still to rate as you work through them. Please don't be alarmed that there are 180! Each should only take a few seconds to rate.

Many thanks again for your time in supporting my research.

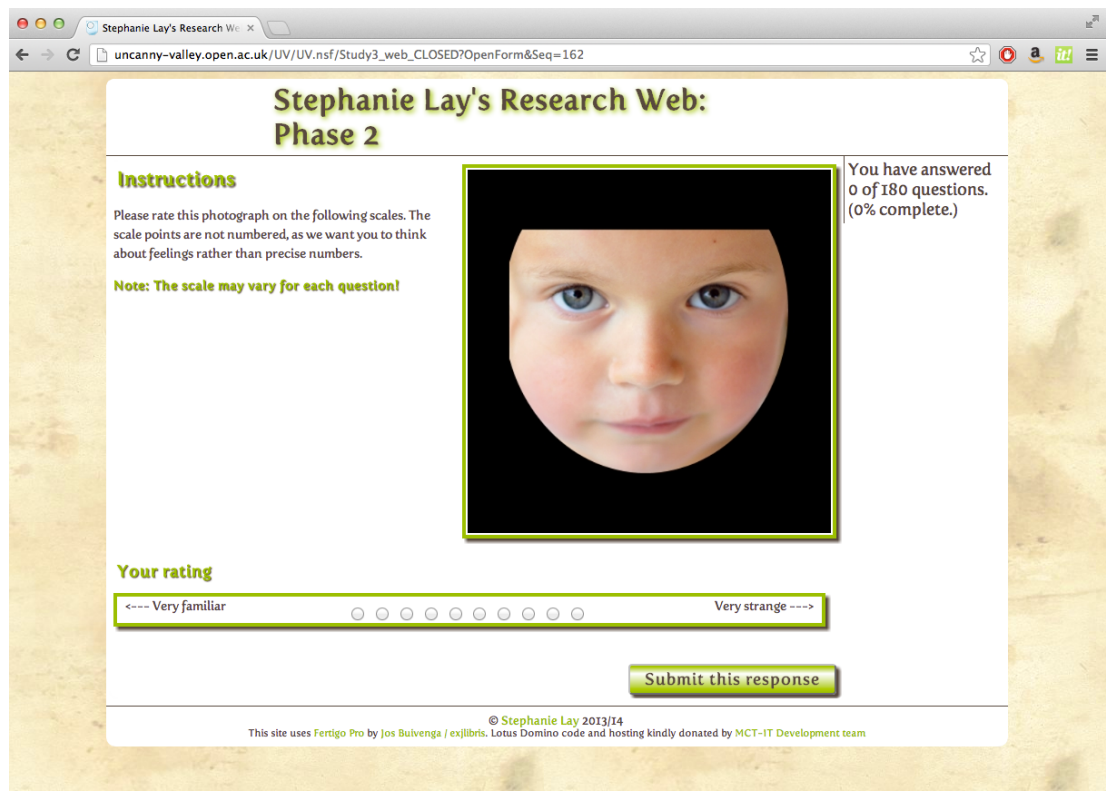


The image below is an example of one of the 180 screens. The image and scale varied to present all 180 combinations of the images and scales.

Instructions

Please rate this photograph on the following scales. The scale points are not numbered, as we want you to think about feelings rather than precise numbers.

Note: The scale may vary for each question!



Demographics

We are almost ready to save your answers.

In order to analyse whether the results vary according to participants' backgrounds, I'd like to know a little about you, please. This information will only be used to analyse

answers by group characteristics, never to identify you personally. Please supply the following:

- Your gender:

Male

Female

- Your age (years):
- Your location (City & Country)

One last thing... (optional)

If you found this study interesting, would you like to take part in more online research for Open University psychologists?

We are setting up a 'virtual participation panel' of people to participate in a wide variety of research projects. You can choose how often you take part, and you can leave the panel at any time. If you'd like to learn more, please leave your email address in the box below. Your email address will be used only for panel recruitment, and won't be used to identify the answers that you have just given.

- Enter your email address:

For more online psychology studies, visit OnlinePsychResearch.co.uk.

Stephanie Lay's Research Web: Phase 2

Demographics

We are almost ready to save your answers.

In order to analyse whether the results vary according to participants' backgrounds, I'd like to know a little about you, please. This information will only be used to analyse answers by group characteristics, never to identify you personally. Please supply the following:

- Your gender:
 - ☐ Male
 - ☐ Female
- Your age (years):
- Your location (City & Country):

 Please select a country...

One last thing... (optional)

If you found this study interesting, would you like to take part in more online research for Open University psychologists?

We are setting up a 'virtual participation panel' of people to participate in a wide variety of research projects. You can choose how often you take part, and you can leave the panel at any time. If you'd like to learn more, please leave your email address in the box below. Your email address will be used only for panel recruitment, and won't be used to identify the answers that you have just given.

- Enter your email address:

For more online psychology studies, visit OnlinePsychResearch.co.uk.

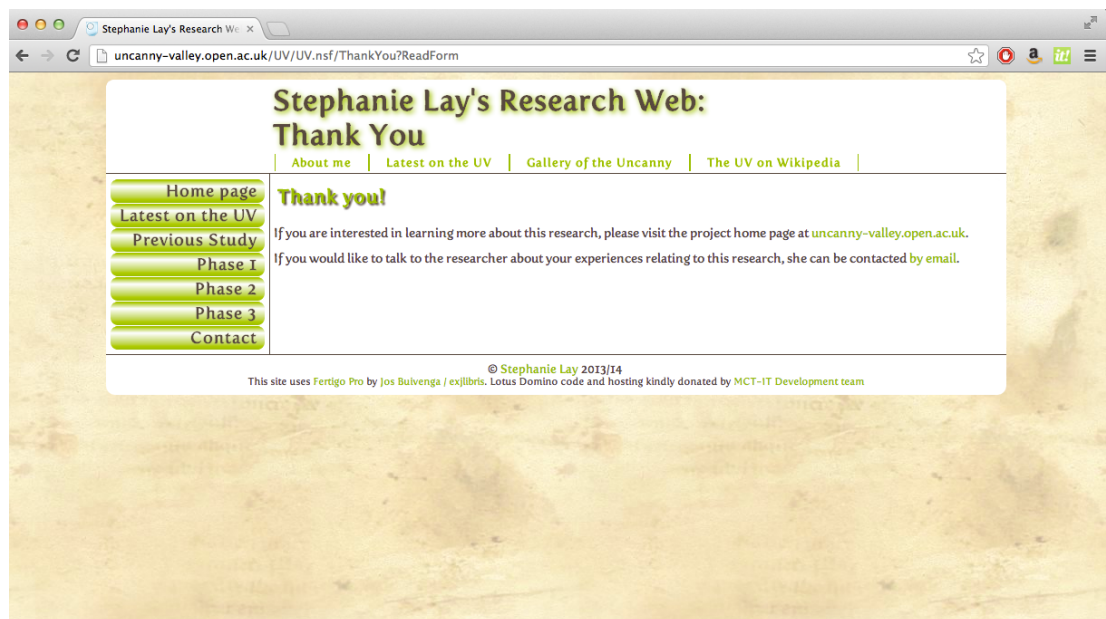
Save all your answers

© Stephanie Lay 2013/I4
 This site uses Fertigo Pro by Jos Bulvenga / exjlibris. Lotus Domino code and hosting kindly donated by MCT-IT Development team

Thank you!

If you are interested in learning more about this research, please visit the project home page at uncanny-valley.open.ac.uk.

If you would like to talk to the researcher about your experiences relating to this research, she can be contacted [by email](#).



Appendix 2f: Participant Information and consent forms for face to face study

Information For Participants

INVITATION TO TAKE PART IN A RESEARCH STUDY

You are being asked to take part in a research study investigating face perception. This research makes up part of my PhD research and is supervised by Dr Nicky Brace and Dr Graham Pike of the Open University, and Dr Frank Pollick of the University of Glasgow.

HOW LONG WILL THE STUDY BE LIVE?

I will be running this experiment from November 19th 2011 until January 19th 2012.

AIM OF THE RESEARCH STUDY

To explore how people perceive faces: humans are particularly adept at recognising faces and the question of why and how they are processed to allow this recognition has been explored in depth by psychologists. You are being invited to take part in a study exploring this mechanism in detail: the results will be used to draw conclusions that will help psychologists understand this important aspect of perception.

WHAT DATA WILL BE COLLECTED?

The study will be made up of blocks of activities. In each activity, you will be shown a picture of a face and asked to remember it. You will then be shown that image along with several similar ones and asked to pick out the one you had learned. I will collect information on whether you pick out the face you had learned, and the time it takes for you to decide on a particular face.

At the end of the study I will ask for some basic demographic information so I can make sure that my conclusions are fair and not biased towards or against any particular group. This part is optional: you can still take part and choose not to give this information.

I may also collect your email address at the end of the study if you wish to join my mailing list to receive a short report of the results of the study.

HOW WILL THE DATA BE COLLECTED?

The images will be shown on a laptop using a software package that will record your answers. You will use the keyboard to indicate your choice of image and move between the blocks of the experiment.

CONFIDENTIALITY/ANONYMITY

The data I collect do not contain any personally sensitive information. If you choose to give any demographic information this will not be linked to the ratings you give in the main part of the activity.

This research is carried out in compliance with the Data Protection and Freedom of Information Acts.

TIME COMMITMENT

The study has been tested and will require around ten minutes to complete: this will take place in a single session. This is an approximate time as there are breaks built into the experiment so you can work at your own pace: you may find you complete more quickly than that, or take a little longer.

WITHDRAWING FROM THE STUDY

You do not have to complete the study: you have the right to withdraw at any time, and you do not have to give me a reason for withdrawing. If you decide you do not want to continue, just let me know and I will halt the session.

Your data will be securely deleted from the laptop, and will not be used in the final analysis of the results.

If you complete the experiment and decide at a later date that you would like to withdraw your data, please contact me before the study closing date and I will ensure they are securely deleted.

RISKS

There are no known risks for you in taking part in this activity.

COST, REIMBURSEMENT AND COMPENSATION

Your participation in this study is voluntary and no reimbursement or compensation is offered for your time. However, you will be contributing to research in a novel area of psychology, and your participation is greatly appreciated.

FURTHER INFORMATION ABOUT THIS RESEARCH PROJECT

When you have completed the study I will be able to tell you more about my overall research project and how your participation in this study has contribute to the broader area of psychology.

There will also be an opportunity to leave your email address if you wish to receive a short report on the results of this study.

CONTACTS:

I (Stephanie Lay) will be glad to answer your questions about this study at any time. If you have a query that I am unable to answer, please contact one of my supervisors:

| Principle Investigator | Supervisor | Supervisor |
|--|--|--|
| Stephanie Lay | Nicky Brace | Graham Pike |
| Psychology Department | Psychology Department | Psychology Department |
| The Open University | The Open University | The Open University |
| Walton Hall | Walton Hall | Walton Hall |
| Milton Keynes | Milton Keynes | Milton Keynes |
| MK7 6AA | MK7 6AA | MK7 6AA |
| s.c.lay@open.ac.uk | n.a.brace@open.ac.uk | g.e.pike@open.ac.uk |

Participation consent form

Agreement to Participate, The Faculty of Social Sciences, The Open University

Agreement to Participate: Exploring Face Perception

I, (print name) agree to take part in this research project.

I have had the purposes of the research project explained to me.

I have been informed that I may refuse to participate at any point by simply saying so.

I have been assured that my confidentiality will be protected as specified in the information sheet.

I agree that the information that I provide can be used for research purposes, including publication.

I understand that if I have any concerns or difficulties I can contact Stephanie Lay, the principle investigator at:

| |
|--|
| Stephanie Lay |
| Psychology Department |
| The Open University |
| Walton Hall |
| Milton Keynes |
| MK7 6AA |
| s.c.lay@open.ac.uk |

If I want to talk to someone else about this project, I can contact Stephanie's supervisors at:

| | |
|--|--|
| Nicky Brace | Graham Pike |
| Psychology Department | Psychology Department |
| The Open University | The Open University |
| Walton Hall | Walton Hall |
| Milton Keynes | Milton Keynes |
| MK7 6AA | MK7 6AA |
| n.a.brace@open.ac.uk | g.e.pike@open.ac.uk |

I assign the copyright for my contribution to the Faculty for use in research and publication.

Signed: (signature)

| | |
|--------------------|--|
| For admin use only | |
| Participant: | |
| Condition: | |

Date: (date)

Appendix 2g: Screenshots from face to face study

Screenshots of Superlab stimulus presentation software screens.

NB: The full text is presented as well as images of the pages.

Hello, and welcome to the experiment. I'm interested in how people recognise faces.

I'm going to show you a series of different faces, one at a time, and ask you to try and

remember each one. When you've had some time to look at each face I'll ask you to pick out the face you just saw from a set of three. Sometimes the faces will be shown upside down, sometimes they will be the right way up. You'll get a chance to practice before we start the live test, and you'll get a break between after you've picked out each picture: please feel free to take your time.

Do you have any questions? If so, ask me now.

If not, we'll start with a practice run. Press any key to get started.

Hello, and welcome to the experiment.

I'm interested in how people recognise faces. I'm going to show you a series of different faces, one at a time, and ask you to try and remember each one.

When you've had some time to look at each face I'll ask you to pick out the face you just saw from a set of three.

Sometimes the faces will be shown upside down, sometimes they will be the right way up.

You'll get a chance to practice before we start the live test, and you'll get a break between after you've picked out each picture: please feel free to take your time.

Do you have any questions? If so, ask me now.

If not, we'll start with a practice run. Press any key to get started.

PRACTICE RUN

Here's a practice run to help you get used to the controls.

The next screen will show you a piece of fruit for three seconds.

I'd like you to look at it carefully: in a moment you will be asked to pick it out of a 'line up' including two other pieces of fruit.

On the laptop keyboard, there are three keys marked out for you to use to indicate the position of the piece of fruit you recognise as the one you just saw:

'L' key - On the left

'M' key - In the middle

'R' key - On the right

Press any key to continue...

PRACTICE RUN

Here's a practice run to help you get used to the controls.

The next screen will show you a piece of fruit for three seconds.

I'd like you to look at it carefully: in a moment you will be asked to pick it out of a 'line up' including two other pieces of fruit.

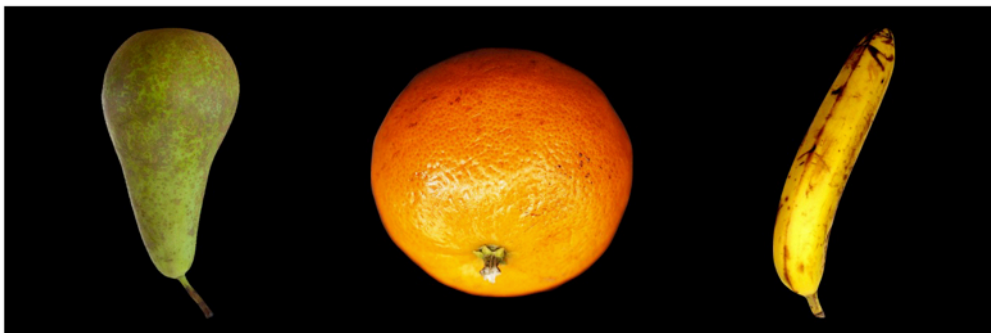
On the laptop keyboard, there are three keys marked out for you to use to indicate the position of the piece of fruit you recognise as the one you just saw:

"L" key - On the left

"M" key - In the middle

"R" key - On the right

Press any key to continue...

**LIVE RUN**

That completes your trial run. The next images you see will be faces, and there will be 24 to view. As before, you will see the face for three seconds. I'd like you to look at it carefully: after three seconds you will be asked to pick it out of a 'line up' including two other faces.

Use the keys to indicate the position of the face you saw:

'L' key - On the left

'M' key - In the middle

'R' key - On the right

Press any key to continue...

PRACTICE RUN

That completes your trial run. The next images you see will be faces, and there will be 24 to view.

As before, you will see the face for three seconds.

I'd like you to look at it carefully: after three seconds you will be asked to pick it out of a 'line up' including two other faces.

Use the keys to indicate the position of the face you saw:

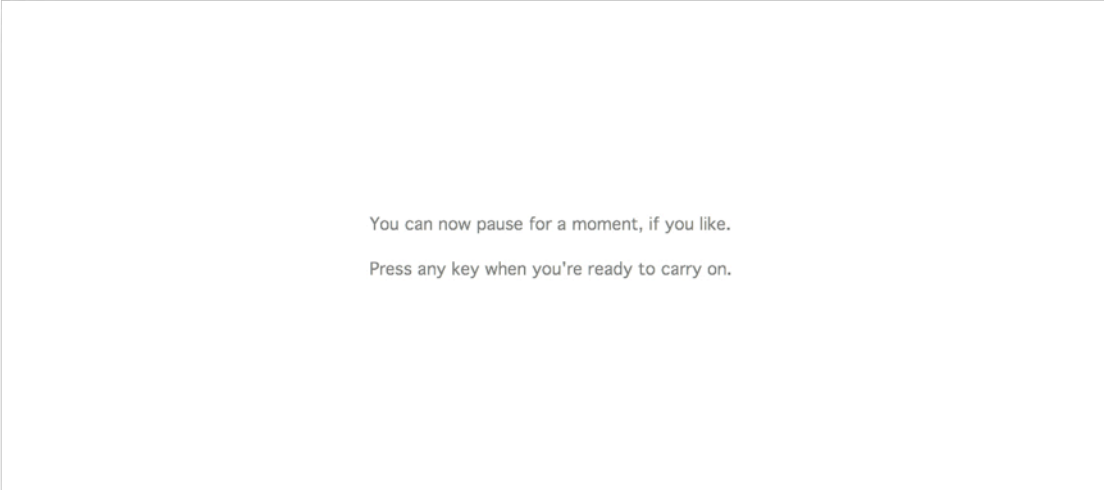
"L" key - On the left

"M" key - In the middle

"R" key - On the right

Press any key to continue...





You can now pause for a moment, if you like.
Press any key when you're ready to carry on.

NB: This sequence of a target image followed by a testing panel and then a break was repeated another 23 times to present all possible images from that morph stage condition.

Thank you! I've now collected all the information that you need to provide.

I have a theory that human faces and almost human faces are processed in a different way and the information that you've provided will allow me to see if there is a difference in how faces are processed.

The types of faces you saw were randomly chosen. They were either human faces, not human at all (statue, dolls, robots or animals) or somewhere in-between where I've taken human faces and 'morphed' them using a computer program so they are somewhere between human and not human.

Sometimes you saw the faces the right way up, and sometimes they were inverted.

I will compare how you all recognised the faces to see which groups were easiest and hardest to recognise.

I would be grateful if you could keep this explanation to yourself if you speak to anyone else who may take part - it's important that the purpose of the experiment isn't known beforehand, as this could alter the way people respond.

If you'd like to learn more about my research, please leave me your email address. For your reference, here are my contact details:

Stephanie Lay,
Psychology Department
The Open University

Phase 3

Appendix 3a: Participant demographic tables

The 1077 participants were asked if they were willing to give demographic information. 999 were willing to give that information. The table below summarises their demographics.

| Sex | N | % |
|----------------|------------|--------------|
| Female | 660 | 66.1 |
| Male | 328 | 32.8 |
| Did not answer | 11 | 1.1 |
| Total | 999 | 100.0 |
| | | |
| Age | N | % |
| < 21 | 50 | 5.0 |
| 21- 25 | 124 | 12.4 |
| 26 - 30 | 168 | 16.8 |
| 31 - 35 | 166 | 16.6 |
| 36 - 40 | 130 | 13.0 |
| 41 - 45 | 112 | 11.2 |
| 46 - 50 | 95 | 9.5 |
| 51 - 55 | 66 | 6.6 |
| 56 - 60 | 62 | 6.2 |
| >60 | 26 | 2.6 |
| Total | 999 | 100.0 |
| | | |
| Country | N | % |
| Afghanistan | 1 | 0.1 |
| Argentina | 1 | 0.1 |
| Australia | 31 | 3.1 |
| Austria | 3 | 0.3 |
| Bangladesh | 1 | 0.1 |
| Belgium | 4 | 0.4 |
| Brazil | 2 | 0.2 |
| Bulgaria | 2 | 0.2 |
| Canada | 39 | 3.9 |
| Colombia | 2 | 0.2 |
| Cyprus | 1 | 0.1 |
| Czech Republic | 1 | 0.1 |

| | | |
|--|------------|--------------|
| Denmark | 1 | 0.1 |
| Ecuador | 1 | 0.1 |
| Finland | 10 | 1.0 |
| France | 6 | 0.6 |
| Germany | 10 | 1.0 |
| Iceland | 2 | 0.2 |
| India | 3 | 0.3 |
| Ireland | 7 | 0.7 |
| Italy | 2 | 0.2 |
| Japan | 1 | 0.1 |
| Latvia | 1 | 0.1 |
| Lithuania | 2 | 0.2 |
| Mexico | 1 | 0.1 |
| Netherlands | 9 | 0.9 |
| New Zealand | 6 | 0.6 |
| Norway | 3 | 0.3 |
| Poland | 2 | 0.2 |
| Portugal | 2 | 0.2 |
| Russian Federation | 2 | 0.2 |
| Saudi Arabia | 1 | 0.1 |
| South Africa | 2 | 0.2 |
| Spain | 4 | 0.4 |
| Switzerland | 2 | 0.2 |
| Turkey | 1 | 0.1 |
| United Kingdom of Great Britain and Northern Ireland | 472 | 47.2 |
| United States of America | 353 | 35.3 |
| Uruguay | 1 | 0.1 |
| Did not answer | 4 | 0.4 |
| Total | 999 | 100.0 |
| | | |
| Ethnicity | N | % |
| African | 1 | 0.1 |
| Any other Asian background | 6 | 0.6 |
| Any other Black/African/Caribbean background | 2 | 0.2 |
| Any other ethnic group | 16 | 1.6 |
| Any other Mixed/multiple ethnic background | 37 | 3.7 |
| Any other White background | 306 | 30.6 |
| Bangladeshi | 1 | 0.1 |

| | | |
|---|------------|--------------|
| Caribbean | 3 | 0.3 |
| Chinese | 4 | 0.4 |
| English / Welsh / Scottish / Northern Irish / British | 558 | 55.9 |
| Gypsy or Irish Traveller | 2 | 0.2 |
| Indian | 8 | 0.8 |
| Irish | 24 | 2.4 |
| Pakistani | 2 | 0.2 |
| White and Asian | 16 | 1.6 |
| White and Black African | 4 | 0.4 |
| White and Black Caribbean | 3 | 0.3 |
| Did not answer | 6 | 0.6 |
| Total | 999 | 100.0 |

Table 31: Participant demographics for Phase 3

Appendix 3b: Survey web page examples

The following images present the survey screens used to collect participant responses.



Next >>

Face Perception Online Study

Thank you for your interest in my study. You are being invited to take part in a research study investigating face perception. This research makes up part of my PhD research and is supervised by Dr Nicky Brace and Professor Graham Pike of the Open University, and Dr Frank Pollick of the University of Glasgow.

Who can take part in this study?
Anyone over the age of 18 can take part.

What happens as part of the study?
This study involves looking at images of faces online and either answering questions about them or using them in a drag and drop activity: this is to investigate psychological aspects of face perception.

What information will be collected?
I will collect the answers that you give to each question and where you place them in the drag and drop activity. I will also ask you for some basic demographic information. This part is optional: you can still take part and choose not to give this information. If you would like to join my mailing list to receive a short report about my findings you can leave your email address at the end of the study.

Confidentiality/Anonymity
The research data I collect do not contain any personally sensitive information. This research is carried out in compliance with the Data Protection and Freedom of Information Acts.

How long will the study take?
The study has been tested and will require around 15 minutes to complete: you can complete the study in a single session, but if you want to, you can stop at any point and come back to finish the study when you are ready. You will need to use the same computer and web browser to carry on where you left off.

Withdrawing from the study
You may decide to stop taking part in the study at any time without explanation. If you decide you no longer want to carry on, just close your web browser. Your responses up to that point will be excluded from the final analysis. If you complete the study and decide at a later date that you would like to withdraw your responses, please contact me before the study closing date and I will ensure they are securely deleted.

Risks
There are no known risks for you in taking part in this study.

Cost, reimbursement and compensation
Your participation in this study is voluntary and no reimbursement or compensation is offered for your time. However, you will be contributing to research in a novel area of psychology, and your participation is greatly appreciated.

Further information about this project
After the questions and activities you will be able to read more about this research and how your participation in this study has contributed to the overall project. There will also be the opportunity to leave your email address, which will not be linked to or stored with the research data collected, but will allow you to receive a short report on the results once the study has been finished.

How long will the study be live?
The study will run from October 29th to November 30th 2012.

Contact information
I (Stephanie Lay) will be glad to answer your questions about this study at any time. If you have a query that I am unable to answer, please contact one of my supervisors: Nicky Brace (nicola.brace@open.ac.uk) or Graham Pike (graham.pike@open.ac.uk).

Please select the option below that applies to you:

☐ I am 18 years old or older and give my consent to take part in this study

☐ I am under 18 years old or do not wish to take part in this study

Next >>

[Next >>](#)

Thank you for giving your consent to take part in this study. As described, you will see several pictures of faces and be asked some questions about them.

There are two parts to the study.

In the first part you will see a series of twenty-four pictures of different faces. Some of the faces will be upside down and some will be the right way up. The face will be presented one at a time and there will be three questions to answer about each face.

In the second part you will see a 'stack' of pictures of twelve faces: your task is to put them into groups depending on the kind of expression you think they are showing.


You will have a chance to practice before each part before you start giving your answer for real so you can see the type of images you will be looking at, view the questions and practice the task.

The study has been tested on several different web browsers: I recommend Chrome, Firefox or Safari but Internet Explorer will work too. While it is possible to complete it using a mobile phone I recommend a desktop, laptop or tablet computer as some of the images are hard to see on a phone.

If you have any problems at any point during the experiment, please contact me at stephanie.lay@open.ac.uk

[Next >>](#)

Task 1: Q1 - Emotion Strength




Picture 1/24

Q1. When looking at the person in the picture, how strongly do you feel each of the emotions below?

| | I don't feel this at all | | | | | I feel this moderately strongly | | | | | I feel this extremely strongly |
|-----------|--------------------------|-----------------------|-----------------------|-----------------------|-----------------------|---------------------------------|-----------------------|-----------------------|-----------------------|-----------------------|--------------------------------|
| Angry | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Disgusted | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Afraid | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Happy | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Sad | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Surprised | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

Next >>

Task 1 - Q2 - Eeriness



Picture 1/24

Q2. How eerie do you find the person in the picture?

Not at all eerie

☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

Extremely eerie

<< Next >>

Task 1 - Q3 - Which emotion is being displayed?



Picture 1/24

Q3. Which of the following best describes the emotion shown by the person in the picture?

- ☐ Anger
- ☐ Disgust
- ☐ Fear
- ☐ Happiness
- ☐ Sadness
- ☐ Surprise
- ☐ No emotion

<< Next >>

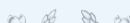
Task 2 - Categorising and Ranking Images

Which type of emotion is being shown by the person in each picture?

Please drag and drop these pictures (on the left) into one of the categories (on the right). The pictures initially appear on top of each other, so that when you move one the next one will be visible. You can use each of the categories as many times as you like and indeed do not need to use all the categories. The layout of the categories will update as you add items to use as much of the screen space as possible.

| Items | Positive emotion | Neutral emotion | Negative emotion |
|-------|------------------|-----------------|------------------|
| | | | |
| | I can't decide | | |

<< Next >>



Demographic data collection screens

Demographic Information

In order to analyse whether the results vary according to participants' backgrounds, I'd like to know a little about you. This information will only be used to analyse answers by group characteristics, never to identify you personally.

The demographic questions ask for sex, age, location and ethnicity but if you prefer not to give these, please select the option and they will not be presented.

☐ I would prefer not to give any demographic data.

Next >>

Your sex:

☐ Male

☐ Female

Your age (in years)

Your location: City

Your location: Country

Which of the following best describes your ethnic group or background?

☐ English / Welsh / Scottish / Northern Irish / British

☐ Irish

☐ Gypsy or Irish Traveller

☐ Any other White background

☐ White and Black Caribbean

☐ White and Black African

☐ White and Asian

☐ Any other Mixed/multiple ethnic background

☐ Indian

☐ Pakistani

☐ Bangladeshi

☐ Chinese

☐ Any other Asian background

☐ African

☐ Caribbean

☐ Any other Black/African/Caribbean background

☐ Arab

☐ Any other ethnic group

Please give your email address if you would like to receive a short report on the results of this study.

<<

Next >>

All Phases

Appendix 4: Eeriness Ratings across all Phases

| Phase | Face description | Orientation | Mean Eeriness |
|-------|--|-------------|---------------|
| 1 | Cluster 1 Face (Low human, highly strange) | Upright | 9.6 |
| 1 | Cluster 2 Face (Medium human-likeness and strangeness) | Upright | 7.0 |
| 1 | Non-human anchor (robot) | Upright | 3.5 |
| 1 | Human anchor | Upright | 2.9 |
| 1 | Cluster 3 Face (High human, low strange and eerie) | Upright | 2.4 |
| 2 | Animal - 50% human | Upright | 7.7 |
| 2 | Robot - 50% human | Upright | 6.8 |
| 2 | Robot - 75% human | Upright | 6.4 |
| 2 | Robot - 25% human | Upright | 6.3 |
| 2 | Animal - 75% human | Upright | 6.1 |
| 2 | Statue - 50% human | Upright | 6.0 |
| 2 | Statue - 25% human | Upright | 6.0 |
| 2 | Doll - Non-human | Upright | 5.8 |
| 2 | Statue - Non-human | Upright | 5.7 |
| 2 | Statue - 75% human | Upright | 5.3 |
| 2 | Doll - 25% human | Upright | 5.1 |
| 2 | Animal - 25% human | Upright | 5.0 |
| 2 | Doll - 50% human | Upright | 4.6 |
| 2 | Robot - Non-human | Upright | 4.6 |
| 2 | Statue - Human | Upright | 3.4 |
| 2 | Doll 75% human | Upright | 3.2 |
| 2 | Animal - Human | Upright | 3.0 |
| 2 | Animal - Non-human | Upright | 2.4 |
| 2 | Doll - Human | Upright | 2.4 |
| 2 | Robot - Human | Upright | 2.3 |
| 3 | Happy face, fearful eyes | Inverted | 5.3 |
| 3 | Happy face, angry eyes | Inverted | 4.9 |
| 3 | Happy face, fearful eyes | Upright | 4.5 |
| 3 | Happy face, sad eyes | Inverted | 3.9 |
| 3 | Happy face, neutral eyes | Inverted | 3.9 |
| 3 | Happy face, angry eyes | Upright | 3.7 |
| 3 | Angry face, happy eyes | Inverted | 3.7 |
| 3 | Disgusted face, neutral eyes | Inverted | 3.5 |
| 3 | Neutral face, fearful eyes | Inverted | 3.5 |
| 3 | Angry face and eyes | Inverted | 3.4 |
| 3 | Angry face, neutral eyes | Inverted | 3.3 |
| 3 | Fearful face and eyes | Inverted | 3.3 |
| 3 | Happy face, sad eyes | Upright | 3.2 |
| 3 | Disgusted face, happy eyes | Inverted | 3.2 |
| 3 | Angry face, happy eyes | Upright | 3.2 |
| 3 | Disgusted face and eyes | Inverted | 3.2 |
| 3 | Neutral face, disgusted eyes | Inverted | 3.1 |
| 3 | Fearful face, happy eyes | Inverted | 3.0 |
| 3 | Sad face, happy eyes | Inverted | 3.0 |
| 3 | Happy face, neutral eyes | Upright | 2.9 |
| 3 | Angry face, neutral eyes | Upright | 2.9 |
| 3 | Sad face and eyes | Inverted | 2.8 |

| Phase | Face description | Orientation | Mean Eeriness |
|-------|------------------------------|-------------|---------------|
| 3 | Neutral face, fearful eyes | Upright | 2.8 |
| 3 | Fearful face, happy eyes | Upright | 2.7 |
| 3 | Fearful face, neutral eyes | Inverted | 2.7 |
| 3 | Neutral face, angry eyes | Inverted | 2.7 |
| 3 | Neutral face, sad eyes | Inverted | 2.7 |
| 3 | Fearful face and eyes | Upright | 2.6 |
| 3 | Angry face and eyes | Upright | 2.6 |
| 3 | Sad face, neutral eyes | Inverted | 2.6 |
| 3 | Disgusted face, neutral eyes | Upright | 2.5 |
| 3 | Happy face and eyes | Inverted | 2.5 |
| 3 | Neutral face, disgusted eyes | Upright | 2.5 |
| 3 | Sad face, happy eyes | Upright | 2.5 |
| 3 | Neutral face and eyes | Inverted | 2.5 |
| 3 | Fearful face, neutral eyes | Upright | 2.4 |
| 3 | Disgusted face, happy eyes | Upright | 2.3 |
| 3 | Disgusted face and eyes | Upright | 2.3 |
| 3 | Neutral face, happy eyes | Inverted | 2.3 |
| 3 | Happy face, disgusted eyes | Inverted | 2.3 |
| 3 | Neutral face, happy eyes | Upright | 2.2 |
| 3 | Happy face and eyes | Upright | 2.1 |
| 3 | Happy face, disgusted eyes | Upright | 2.0 |
| 3 | Neutral face, angry eyes | Upright | 2.0 |
| 3 | Sad face, neutral eyes | Upright | 2.0 |
| 3 | Neutral face, sad eyes | Upright | 2.0 |
| 3 | Neutral face and eyes | Upright | 1.9 |
| 3 | Sad face and eyes | Upright | 1.7 |

Table 32: Table comparing mean eeriness ratings for images used in each research phase.